

P

Access DB# 48171

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: \_\_\_\_\_ Examiner #: \_\_\_\_\_ Date: \_\_\_\_\_  
Art Unit: \_\_\_\_\_ Phone Number 30 \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Mail Box and Bldg/Room Location: \_\_\_\_\_ Results Format Preferred (circle): PAPER DISK E-MAIL

**If more than one search is submitted, please prioritize searches in order of need.**

\*\*\*\*\*

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

### FOR OFFICIAL USE ONLY

#### STAFF USE ONLY

Searcher: B. Kessner - Italy  
Searcher Phone #: 308-4501  
Searcher Location: Bited Lab  
Date Searcher Picked Up: 8/2/01  
Date Completed: 8/6/01  
Searcher Prep & Review Time: 5 mi  
Clerical Prep Time: 4 mi  
Online Time: 2 mi

#### Type of Search

NA Sequence (#) 2  
AA Sequence (#) \_\_\_\_\_  
Structure (#) \_\_\_\_\_  
Bibliographic \_\_\_\_\_  
Litigation \_\_\_\_\_  
Fulltext \_\_\_\_\_  
Patent Family \_\_\_\_\_  
Other \_\_\_\_\_

#### Vendors and cost where applicable

STN \_\_\_\_\_  
Dialog \_\_\_\_\_  
Questel/Orbit \_\_\_\_\_  
Dr. Link \_\_\_\_\_  
Lexis/Nexis \_\_\_\_\_  
Sequence Systems CompuLink 1  
WWW/Internet \_\_\_\_\_  
Other (specify) \_\_\_\_\_

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:57:28 : Search time 183.37 Seconds  
(without alignments)  
178.060 Million cell updates/sec

Title: US-09-701-626A-79

Perfect score: 52

Sequence: 1 tcgcttcgctcactgcggga.....ggcccttaacaaacgttt 52

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601:\*

1: /cgnl\_9/gcgdata/geneseq/geneseq/NA1980.DAT:\*

2: /cgnl\_9/gcgdata/geneseq/geneseq/NA1981.DAT:\*

3: /cgnl\_9/gcgdata/geneseq/geneseq/NA1982.DAT:\*

4: /cgnl\_9/gcgdata/geneseq/geneseq/NA1983.DAT:\*

5: /cgnl\_9/gcgdata/geneseq/geneseq/NA1984.DAT:\*

6: /cgnl\_9/gcgdata/geneseq/geneseq/NA1985.DAT:\*

7: /cgnl\_9/gcgdata/geneseq/geneseq/NA1986.DAT:\*

8: /cgnl\_9/gcgdata/geneseq/geneseq/NA1987.DAT:\*

9: /cgnl\_9/gcgdata/geneseq/geneseq/NA1988.DAT:\*

10: /cgnl\_9/gcgdata/geneseq/geneseq/NA1989.DAT:\*

11: /cgnl\_9/gcgdata/geneseq/geneseq/NA1990.DAT:\*

12: /cgnl\_9/gcgdata/geneseq/geneseq/NA1991.DAT:\*

13: /cgnl\_9/gcgdata/geneseq/geneseq/NA1992.DAT:\*

14: /cgnl\_9/gcgdata/geneseq/geneseq/NA1993.DAT:\*

15: /cgnl\_9/gcgdata/geneseq/geneseq/NA1994.DAT:\*

16: /cgnl\_9/gcgdata/geneseq/geneseq/NA1995.DAT:\*

17: /cgnl\_9/gcgdata/geneseq/geneseq/NA1996.DAT:\*

18: /cgnl\_9/gcgdata/geneseq/geneseq/NA1997.DAT:\*

19: /cgnl\_9/gcgdata/geneseq/geneseq/NA1998.DAT:\*

20: /cgnl\_9/gcgdata/geneseq/geneseq/NA1999.DAT:\*

21: /cgnl\_9/gcgdata/geneseq/geneseq/NA2000.DAT:\*

22: /cgnl\_9/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	52	21	AAZ44968
2	52	100.0	75	21	AAZ88508
3	52	100.0	78	21	AAZ44912
4	52	100.0	78	21	AAZ44927
5	52	100.0	78	21	AAZ44936
6	52	100.0	78	21	AAZ44938
7	52	100.0	78	21	AAZ44940
8	52	100.0	78	21	AAZ44948
9	52	100.0	7300	21	AAZ44982
10	52	100.0	7300	21	AAZ44983
11	52	100.0	13382	21	AAZ44984

12	52	100.0	14143	21	AAZ44981	P. alcaligenes res
13	50.4	96.9	75	21	AAZ88514	P. alcaligenes rep
14	50.4	96.9	78	21	AAZ44956	P. alcaligenes rep
15	49.4	95.0	79	21	AAZ44941	P. alcaligenes rep
16	40	76.9	74	21	AAZ88505	P. alcaligenes rep
17	40	76.9	74	21	AAZ88511	P. alcaligenes rep
18	40	76.9	74	21	AAZ88515	P. alcaligenes rep
19	40	76.9	74	21	AAZ88516	P. alcaligenes rep
20	40	76.9	74	21	AAZ88520	P. alcaligenes rep
21	40	76.9	77	21	AAZ44900	P. alcaligenes rep
22	40	76.9	77	21	AAZ44908	P. alcaligenes rep
23	40	76.9	77	21	AAZ44911	P. alcaligenes rep
24	40	76.9	77	21	AAZ44913	P. alcaligenes rep
25	40	76.9	77	21	AAZ44930	P. alcaligenes rep
26	40	76.9	77	21	AAZ44944	P. alcaligenes rep
27	40	76.9	77	21	AAZ44952	P. alcaligenes rep
28	40	76.9	77	21	AAZ44954	P. alcaligenes rep
29	40	76.9	77	21	AAZ44957	P. alcaligenes rep
30	40	76.9	77	21	AAZ44966	P. alcaligenes rep
31	40	76.9	77	21	AAZ44967	P. alcaligenes rep
32	40	76.9	77	21	AAZ44987	P. alcaligenes rep
33	40	76.9	78	21	AAZ44907	P. alcaligenes rep
34	40	76.9	78	21	AAZ44933	P. alcaligenes rep
35	40	76.9	78	21	AAZ44997	P. alcaligenes rep
36	38.4	73.8	74	21	AAZ88503	P. alcaligenes rep
37	38.4	73.8	77	21	AAZ44894	P. alcaligenes rep
38	38.4	73.8	77	21	AAZ44910	P. alcaligenes rep
39	38.4	73.8	77	21	AAZ44915	P. alcaligenes rep
40	38.4	73.8	77	21	AAZ44929	P. alcaligenes rep
41	38.4	73.8	77	21	AAZ44964	P. alcaligenes rep
42	38.4	73.8	77	21	AAZ44990	P. alcaligenes rep
43	38	73.1	76	21	AAZ44961	P. alcaligenes rep
44	37.4	71.9	77	21	AAZ44937	P. alcaligenes rep
45	37	71.2	79	21	AAZ44986	P. alcaligenes rep

#### ALIGNMENTS

RESULT 1  
AAZ44968  
ID AAZ44968 standard; DNA; 52 BP.  
AC AAZ44968;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes repeat (PAR) element DNA #75.  
DE Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; repeat element; PAR; ss.  
KW  
XX  
OS Pseudomonas alcaligenes.  
XX  
XX WO9964632-AL.  
PN  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
XX (NEWE ) NEW ENGLAND BIOLABS INC.  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
XX WPI; 2000-116558/10.  
XX  
PT Cloning intact genes used to isolate genes for restriction enzymes  
XX Claim 10; Page 62; 97pp; English.  
PS  
XX

CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites,  
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
 CC enzymes; toxins that interact with a host, e.g. for synthesis of  
 CC inhibitors or antagonists of the toxin, or in vaccination, or a  
 CC modification methyltransferase. Intact genes can be cloned directly with  
 CC a high probability that the orientation of expression is known in advance  
 CC and low probability of association with extraneous, possibly toxic,  
 CC genes. AA24894-244980 represent the *Pseudomonas* alcaligenes repeat (PAK)  
 CC elements described in the method of the invention.

XX Sequence 52 BP; 11 A; 19 C; 12 G; 10 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-11;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgttcgcctcactcgggaccggtctaaagcgcgcgccttaaccacaaagtta 52  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 tcgttcgcctcactcgggaccggtctaaagcgcgcgccttaaccacaaagtta 52

RESULT 2  
 AA288508  
 ID AA288508 standard; DNA: 75 BP.  
 AC AA288508;  
 XX  
 DT 16-MAY-2000 (first entry)  
 XX  
 DE P. alcaligenes repeat (PAR) element DNA PAR15.  
 XX  
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; repeat element; PAK; ss.  
 XX  
 OS *Pseudomonas* alcaligenes.  
 XX  
 PN WO9964632-A1.  
 PD 16-DEC-1999.  
 XX  
 PF 11-JUN-1999; 99WO-US13295.  
 XX  
 PR 12-JUN-1998; 98US-0089086.  
 PR 12-JUN-1998; 98US-0089101.  
 XX  
 PA (NEW) NEW ENGLAND BIOLABS INC.  
 XX  
 PI Raleigh EA, Vaisville R, Morgan RD.  
 XX  
 DR WP1; 2000-116558/10.  
 XX  
 PT Cloning intact genes used to isolate genes for restriction enzymes -  
 XX Example 1B; Fig 3E; 97pp: English.

CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction

CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites,  
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
 CC enzymes; toxins that interact with a host, e.g. for synthesis of  
 CC inhibitors or antagonists of the toxin, or in vaccination, or a  
 CC modification methyltransferase. Intact genes can be cloned directly with  
 CC a high probability that the orientation of expression is known in advance  
 CC and low probability of association with extraneous, possibly toxic,  
 CC genes. AA288504-288521 represent *Pseudomonas* alcaligenes repeat (PAK)  
 CC elements described in the method of the invention.

XX Sequence 75 BP; 18 A; 23 C; 17 G; 17 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-11;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgttcgcctcactcgggaccggtctaaagcgcgcgccttaaccacaaagtta 52  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 22 tcgttcgcctcactcgggaccggtctaaagcgcgcgccttaaccacaaagtta 73

RESULT 3  
 AA244912  
 ID AA244912 standard; DNA: 78 BP.  
 AC AA244912;  
 XX  
 DT 16-MAY-2000 (first entry)  
 XX  
 DE P. alcaligenes repeat (PAK) element DNA #19.  
 XX  
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; repeat element; PAK; ss.  
 XX  
 OS *Pseudomonas* alcaligenes.  
 XX  
 PN WO9964632-A1.  
 PD 16-DEC-1999.  
 XX  
 PF 11-JUN-1999; 99WO-US13295.  
 XX  
 PR 12-JUN-1998; 98US-0089086.  
 PR 12-JUN-1998; 98US-0089101.  
 XX  
 PA (NEW) NEW ENGLAND BIOLABS INC.  
 XX  
 PI Raleigh EA, Vaisville R, Morgan RD;  
 XX  
 DR WP1; 2000-116558/10.  
 XX  
 PT Cloning intact genes used to isolate genes for restriction enzymes -  
 XX Claim 7a; Page 59; 97pp: English.

CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites,  
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
 CC enzymes; toxins that interact with a host, e.g. for synthesis of  
 CC inhibitors or antagonists of the toxin, or in vaccination, or a  
 CC modification methyltransferase. Intact genes can be cloned directly with  
 CC a high probability that the orientation of expression is known in advance  
 CC and low probability of association with extraneous, possibly toxic,

CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)  
 CC elements described in the method of the invention.

XX Sequence 78 BP; 18 A; 24 C; 19 G; 17 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 78;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-11;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggagccggtctaaagcggcccttaaccacaaagtta 52  
 |||||  
 DB 24 tcgcttcgctcactgcggagccggtctaaagcggcccttaaccacaaagtta 75

RESULT 4  
 AAZ44927  
 ID AAZ44927 standard; DNA; 78 BP.  
 XX  
 AC AAZ44927;  
 XX  
 DT 16-MAY-2000 (first entry)  
 DE  
 XX  
 XX P. alcaligenes repeat (PAR) element DNA #34.  
 XX  
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; repeat element; PAR; ss.  
 XX

OS Pseudomonas alcaligenes.

PN W09964632-A1.

PD 16-DEC-1999.

PF 11-JUN-1999; 99WO-US13295.

PR 12-JUN-1998; 98US-0089086.

PR 12-JUN-1998; 98US-0089101.

XX (NEWE ) NEW ENGLAND BIOLABS INC.

PI Raleigh EA, Vaisvila R, Morgan RD;

XX WPI; 2000-116558/10.

PT Cloning intact genes used to isolate genes for restriction enzymes

PS Claim 7a; Page 59; 97pp; English.

CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites,  
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
 CC enzymes; toxins that interact with a host, e.g. for synthesis of  
 CC inhibitors or antagonists of the toxin, or in vaccination, or a  
 CC modification methyltransferase. Intact genes can be cloned directly with  
 CC a high probability that the orientation of expression is known in advance  
 CC and low probability of association with extraneous, possibly toxic,  
 CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)  
 CC elements described in the method of the invention.

XX Sequence 78 BP; 18 A; 23 C; 18 G; 19 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 78;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-11;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggagccggtctaaagcggcccttaaccacaaagtta 52  
 |||||  
 DB 24 tcgcttcgctcactgcggagccggtctaaagcggcccttaaccacaaagtta 75

RESULT 5  
 AAZ44936  
 ID AAZ44936 standard; DNA; 78 BP.  
 XX  
 AC AAZ44936;  
 XX  
 DT 16-MAY-2000 (first entry)  
 DE  
 XX  
 XX P. alcaligenes repeat (PAR) element DNA #43.  
 XX  
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; repeat element; PAR; ss.

OS Pseudomonas alcaligenes.

PN W09964632-A1.

PD 16-DEC-1999.

PF 11-JUN-1999; 99WO-US13295.

PR 12-JUN-1998; 98US-0089086.

PR 12-JUN-1998; 98US-0089101.

XX (NEWE ) NEW ENGLAND BIOLABS INC.

PI Raleigh EA, Vaisvila R, Morgan RD;

XX WPI; 2000-116558/10.

PT Cloning intact genes used to isolate genes for restriction enzymes

PS Claim 7a; Page 60; 97pp; English.

CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites,  
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
 CC enzymes; toxins that interact with a host, e.g. for synthesis of  
 CC inhibitors or antagonists of the toxin, or in vaccination, or a  
 CC modification methyltransferase. Intact genes can be cloned directly with  
 CC a high probability that the orientation of expression is known in advance  
 CC and low probability of association with extraneous, possibly toxic,  
 CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)  
 CC elements described in the method of the invention.

XX Sequence 78 BP; 17 A; 24 C; 19 G; 18 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 78;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-11;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggagccggtctaaagcggcccttaaccacaaagtta 52  
 |||||  
 DB 24 tcgcttcgctcactgcggagccggtctaaagcggcccttaaccacaaagtta 75

RESULT 6

```
AAZ44938
ID  AAZ44938 standard; DNA: 78 BP.
AC  AAZ44938;
XX
XX  16-MAY-2000 (first entry)
DE
XX  P. alcaligenes repeat (PAR) element DNA #45.
XX
XX  Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW  detoxifying enzyme; repeat element; PAR; ss.
XX
OS  Pseudomonas alcaligenes.
XX
XX  WO9964632-A1.
XX
XX  16-DEC-1999.
XX
XX  11-JUN-1999; 99WO-US13295.
XX
XX  12-JUN-1998; 98US-0089086.
XX
XX  12-JUN-1998; 98US-0089101.
XX
XX  (NEW) NEW ENGLAND BIOLABS INC.
XX
XX  Raleigh EA, Vaisvila R, Morgan RD;
PI  WPI: 2000-116558/10.
XX
XX  Cloning intact genes used to isolate genes for restriction enzymes
PT
XX  Claim 7a; Page 60; 97pp; English.
XX
XX  This invention describes a novel method for cloning intact,
CC  diversity-selected genes (I) from within gene cassettes (GC) which
CC  comprises identifying DNA repeats that flank GC, hybridizing
CC  oligonucleotides (ON) to these repeats and amplification to produce DNA
CC  fragments containing (I), ligating these fragments into a vector and
CC  transforming cells with the vector. This method is used to clone a wide
CC  variety of prokaryotic genes that provide a selective advantage under
CC  particular conditions, particularly those that encode restriction
CC  enzymes (used as reagents in molecular biology); adhesins (for use in
CC  coating or for targeting molecules or organisms to particular sites,
CC  e.g. for competitive exclusion of a selected pathogen); detoxifying
CC  enzymes; toxins that interact with a host, e.g. for synthesis of
CC  inhibitors or antagonists of the toxin, or in vaccination, or a
CC  modification methyltransferase. Intact genes can be cloned directly with
CC  a high probability that the orientation of expression is known in advance
CC  and low probability of association with extraneous, possibly toxic,
CC  genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
CC  elements described in the method of the invention.
XX
XX  Sequence 78 BP; 17 A; 26 C; 18 G; 17 T; 0 other;

Query Match      100.0%; Score 52; DB 21; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.4e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  tcgttcgtcactgcgggaccggttaaagccgccccttaacaaacgtta 52
Db  24  tcgttcgtcactgcgggaccggttaaagccgccccttaacaaacgtta 75

RESULT 7
AAZ44940
ID  AAZ44940 standard; DNA: 78 BP.
XX
XX  AAZ44940;
XX
XX  16-MAY-2000 (first entry)
XX
XX  P. alcaligenes repeat (PAR) element DNA #47.
```

```
XX
KW  Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW  detoxifying enzyme; repeat element; PAR; ss.
XX
XX  Pseudomonas alcaligenes.
XX
XX  WO9964632-A1.
XX
XX  16-DEC-1999.
XX
XX  11-JUN-1999; 99WO-US13295.
XX
XX  12-JUN-1998; 98US-0089086.
XX
XX  12-JUN-1998; 98US-0089101.
XX
XX  (NEW) NEW ENGLAND BIOLABS INC.
XX
XX  Raleigh EA, Vaisvila R, Morgan RD;
PI  WPI: 2000-116558/10.
XX
XX  Cloning intact genes used to isolate genes for restriction enzymes
PT
XX  Claim 7a; Page 60; 97pp; English.
XX
XX  This invention describes a novel method for cloning intact,
CC  diversity-selected genes (I) from within gene cassettes (GC) which
CC  comprises identifying DNA repeats that flank GC, hybridizing
CC  oligonucleotides (ON) to these repeats and amplification to produce DNA
CC  fragments containing (I), ligating these fragments into a vector and
CC  transforming cells with the vector. This method is used to clone a wide
CC  variety of prokaryotic genes that provide a selective advantage under
CC  particular conditions, particularly those that encode restriction
CC  enzymes (used as reagents in molecular biology); adhesins (for use in
CC  coating or for targeting molecules or organisms to particular sites,
CC  e.g. for competitive exclusion of a selected pathogen); detoxifying
CC  enzymes; toxins that interact with a host, e.g. for synthesis of
CC  inhibitors or antagonists of the toxin, or in vaccination, or a
CC  modification methyltransferase. Intact genes can be cloned directly with
CC  a high probability that the orientation of expression is known in advance
CC  and low probability of association with extraneous, possibly toxic,
CC  genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
CC  elements described in the method of the invention.
XX
XX  Sequence 78 BP; 17 A; 25 C; 20 G; 16 T; 0 other;

Query Match      100.0%; Score 52; DB 21; Length 78;
Best Local Similarity 100.0%; Pred. No. 2.4e-11;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  tcgttcgtcactgcgggaccggttaaagccgccccttaacaaacgtta 52
Db  24  tcgttcgtcactgcgggaccggttaaagccgccccttaacaaacgtta 75

RESULT 8
AAZ44948
ID  AAZ44948 standard; DNA: 78 BP.
XX
XX  AAZ44948;
XX
XX  16-MAY-2000 (first entry)
XX
XX  P. alcaligenes repeat (PAR) element DNA #55.
XX
XX  Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW  detoxifying enzyme; repeat element; PAR; ss.
XX
XX  Pseudomonas alcaligenes.
XX
XX  WO9964632-A
```

16-DEC-1999.  
11-JUN-1999; 99WO-US13295.  
12-JUN-1998; 98US-0089086.  
12-JUN-1998; 98US-0089101.  
(NEWE ) NEW ENGLAND BIOLABS INC.  
Raleigh EA, Vaisvila R, Morgan RD;  
WPI: 2000-116558/10.  
Cloning intact genes used to isolate genes for restriction enzymes -  
Claim 7a; Page 60; 97pp; English.  
This invention describes a novel method for cloning intact,  
diversity-selected genes (I) from within gene cassettes (GC) which  
comprises identifying DNA repeats that flank GC, hybridizing  
oligonucleotides (ON) to these repeats and amplification to produce DNA  
fragments containing (I), ligating these fragments into a vector and  
transforming cells with the vector. This method is used to clone a wide  
variety of prokaryotic genes that provide a selective advantage under  
particular conditions, particularly those that encode restriction  
enzymes (used as reagents in molecular biology); adhesins (for use in  
coating or for targeting molecules or organisms to particular sites,  
e.g. for competitive exclusion of a selected pathogen); detoxifying  
enzymes; toxins that interact with a host, e.g. for synthesis of  
inhibitors or antagonists of the toxin, or in vaccination, or a  
modification methyltransferase. Intact genes can be cloned directly with  
a high probability that the orientation of expression is known in advance  
and low probability of association with extraneous, possibly toxic,  
genes. AA244894-244980 represent the Pseudomonas alcaligenes repeat (PAR)  
elements described in the method of the invention.  
Sequence 78 BP; 19 A; 24 C; 18 G; 17 T; 0 other;  
Query Match 100.0%; Score 52; DB 21; Length 78;  
Best Local Similarity 100.0%; Pred. No. 2.4e-11;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tcgcttcgctcactcgcggaccggtctaaagccggcccttaaccaaagctta 52  
|||||  
Db 24 tcgcttcgctcactcgcggaccggtctaaagccggcccttaaccaaagctta 75  
RESULT 9  
AA244982  
ID AA244982 standard; DNA; 7300 BP.  
AC AA244982;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes restriction enzyme PacI superintegron DNA #2.  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; PacI; superintegron; ss.  
XX  
OS Pseudomonas alcaligenes.  
XX  
PN WO9964632-A1.  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
PA (NEWE ) NEW ENGLAND BIOLABS INC.

XX  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
DR WPI: 2000-116558/10.  
XX  
PT Cloning intact genes used to isolate genes for restriction enzymes -  
XX  
PS Example 1A; Fig 3B-1; 97pp; English.  
XX  
CC This invention describes a novel method for cloning intact,  
CC diversity-selected genes (I) from within gene cassettes (GC) which  
CC comprises identifying DNA repeats that flank GC, hybridizing  
CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
CC fragments containing (I), ligating these fragments into a vector and  
CC transforming cells with the vector. This method is used to clone a wide  
CC variety of prokaryotic genes that provide a selective advantage under  
CC particular conditions, particularly those that encode restriction  
CC enzymes (used as reagents in molecular biology); adhesins (for use in  
CC coating or for targeting molecules or organisms to particular sites,  
CC for competitive exclusion of a selected pathogen); detoxifying  
CC toxins that interact with a host, e.g. for synthesis of inhibitors or  
CC antagonists of the toxin, or in vaccination, or a modification  
CC methyltransferase. Intact genes can be cloned directly with a high  
CC probability that the orientation of expression is known in advance and  
CC low probability of association with extraneous, possibly toxic, genes.  
CC This sequence represents a Pseudomonas alcaligenes PacI superintegron  
CC DNA sequence.  
XX  
SQ Sequence 7300 BP; 2023 A; 1944 C; 1641 G; 1690 T; 2 other;  
XX  
Query Match 100.0%; Score 52; DB 21; Length 7300;  
Best Local Similarity 100.0%; Pred. No. 4.7e-11;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 tcgcttcgctcactcgcggaccggtctaaagccggcccttaaccaaagctta 52  
|||||  
Db 3300 tcgcttcgctcactcgcggaccggtctaaagccggcccttaaccaaagctta 3351  
RESULT 10  
AA244983  
ID AA244983 standard; DNA; 7300 BP.  
XX  
AC AA244983;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes restriction enzyme PacI superintegron DNA #3.  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; PacI; superintegron; ss.  
XX  
OS Pseudomonas alcaligenes.  
XX  
PN WO9964632-A1.  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
PA (NEWE ) NEW ENGLAND BIOLABS INC.  
XX  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
DR WPI: 2000-116558/10.  
XX  
PT Cloning intact genes used to isolate genes for restriction enzymes -  
XX  
PS Example 1A; Fig 3C-1; 97pp; English.

XX This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites, e.g.  
 CC for competitive exclusion of a selected pathogen); detoxifying enzymes;  
 CC toxins that interact with a host, e.g. for synthesis of inhibitors or  
 CC antagonists of the toxin, or in vaccination, or a modification  
 CC methyltransferase. Intact genes can be cloned directly with a high  
 CC probability that the orientation of expression is known in advance and  
 CC low probability of association with extraneous, possibly toxic, genes.  
 CC This sequence represents a Pseudomonas alcaligenes PacI superintegron  
 CC DNA sequence.

XX  
 SQ Sequence 7300 BP; 2023 A; 1944 C; 1641 G; 1690 T; 2 other;

Query Match 100.0%; Score 52; DB 21; Length 7300;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-11;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggaccggtctaaagccgccccttaacccaaacgtta 52  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 3300 tcgcttcgctcactgcggaccggtctaaagccgccccttaacccaaacgtta 3351

RESULT 11  
 AA244984  
 ID AA244984 standard; DNA; 13382 BP.  
 XX  
 AC AA244984;  
 XX  
 DT 16-MAY-2000 (first entry)  
 XX  
 DE P. alcaligenes restriction enzyme PacI superintegron DNA #4.  
 XX  
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; PacI; superintegron; ss.  
 XX  
 OS Pseudomonas alcaligenes.  
 XX  
 PN W09964632-AL.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 11-JUN-1999; 99WO-US13295.  
 XX  
 PR 12-JUN-1998; 98US-0089086.  
 PR 12-JUN-1998; 98US-0089101.  
 XX  
 PA (NEW ) NEW ENGLAND BIOLABS INC.  
 XX  
 PI Raleigh EA, Vaisvila R, Morgan RD;  
 XX  
 DR WPI; 2000-116558/10.  
 XX  
 PT Cloning intact genes used to isolate genes for restriction enzymes  
 XX  
 PS Example 1A; Fig 3D-1; 97pp; English.  
 XX  
 CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under

CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites, e.g.  
 CC for competitive exclusion of a selected pathogen); detoxifying enzymes;  
 CC toxins that interact with a host, e.g. for synthesis of inhibitors or  
 CC antagonists of the toxin, or in vaccination, or a modification  
 CC methyltransferase. Intact genes can be cloned directly with a high  
 CC probability that the orientation of expression is known in advance and  
 CC low probability of association with extraneous, possibly toxic, genes.  
 CC This sequence represents a Pseudomonas alcaligenes PacI superintegron  
 CC DNA sequence.

XX  
 SQ Sequence 13382 BP; 3325 A; 3756 C; 3464 G; 2837 T; 0 other;

Query Match 100.0%; Score 52; DB 21; Length 13382;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-11;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcggaccggtctaaagccgccccttaacccaaacgtta 52  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 8270 tcgcttcgctcactgcggaccggtctaaagccgccccttaacccaaacgtta 8321

RESULT 12  
 AA244981  
 ID AA244981 standard; DNA; 14143 BP.  
 XX  
 AC AA244981;  
 XX  
 DT 16-MAY-2000 (first entry)  
 XX  
 DE P. alcaligenes restriction enzyme PacI superintegron DNA #1.  
 XX  
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; PacI; superintegron; ss.  
 XX  
 OS Pseudomonas alcaligenes.  
 XX  
 PN W09964632-AL.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 11-JUN-1999; 99WO-US13295.  
 XX  
 PR 12-JUN-1998; 98US-0089086.  
 PR 12-JUN-1998; 98US-0089101.  
 XX  
 PA (NEW ) NEW ENGLAND BIOLABS INC.  
 XX  
 PI Raleigh EA, Vaisvila R, Morgan RD;  
 XX  
 DR WPI; 2000-116558/10.  
 XX  
 PT Cloning intact genes used to isolate genes for restriction enzymes  
 XX  
 PS Example 1A; Fig 3A-1; 97pp; English.  
 XX  
 CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites, e.g.  
 CC for competitive exclusion of a selected pathogen); detoxifying enzymes;  
 CC toxins that interact with a host, e.g. for synthesis of inhibitors or  
 CC antagonists of the toxin, or in vaccination, or a modification  
 CC methyltransferase. Intact genes can be cloned directly with a high  
 CC probability that the orientation of expression is known in advance and

CC low probability of association with extraneous, possibly toxic, genes.  
 CC This sequence represents a Pseudomonas alcaligenes PacI superintegron  
 CC DNA sequence.

XX  
 SQ Sequence 14143 BP; 3936 A; 3446 C; 3311 G; 3449 T; 1 other;

Query Match 100.0%; Score 52; DB 21; Length 14143;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-11;  
 Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 52  
 |||||  
 Db 4112 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 4163

RESULT 13  
 AA288514  
 ID AA288514 standard; DNA; 75 BP.

AC AA288514;

XX  
 DT 16-MAY-2000 (first entry)

XX P. alcaligenes repeat (PAR) element DNA PARf11.

XX  
 KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; repeat element; PAR; ss.

XX Pseudomonas alcaligenes.

XX WO9964632-A1.

XX 16-DEC-1999.

XX 11-JUN-1999; 99WO-US13295.

XX 12-JUN-1998; 98US-0089086.

XX 12-JUN-1998; 98US-0089101.

XX (NEW) NEW ENGLAND BIOLABS INC.

XX Raleigh EA, Vaisvila R, Morgan RD;

XX WPI; 2000-116558/10.

CC Cloning intact genes used to isolate genes for restriction enzymes -  
 Example 1B; Fig 3E; 97pp; English.

CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites,  
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
 CC enzymes; toxins that interact with a host, e.g. for synthesis of  
 CC inhibitors or antagonists of the toxin, or in vaccination, or a  
 CC modification methyltransferase. Intact genes can be cloned directly with  
 CC a high probability that the orientation of expression is known in advance  
 CC and low probability of association with extraneous, possibly toxic,  
 CC genes. AA288504-288521 represent Pseudomonas alcaligenes repeat (PAR)  
 CC elements described in the method of the invention.

XX Sequence 75 BP; 16 A; 23 C; 19 G; 17 T; 0 other;

Query Match 96.9%; Score 50.4; DB 21; Length 75;

Best Local Similarity 98.1%; Pred. No. 9.9e-11;  
 Matches 51; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 52  
 |||||  
 Db 22 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 73

RESULT 14  
 AA244956  
 ID AA244956 standard; DNA; 78 BP.

XX AA244956;

XX 16-MAY-2000 (first entry)

XX P. alcaligenes repeat (PAR) element DNA #63.

XX Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; repeat element; PAR; ss.

XX Pseudomonas alcaligenes.

XX WO9964632-A1.

XX 16-DEC-1999.

XX 11-JUN-1999; 99WO-US13295.

XX 12-JUN-1998; 98US-0089086.

XX 12-JUN-1998; 98US-0089101.

XX (NEW) NEW ENGLAND BIOLABS INC.

XX Raleigh EA, Vaisvila R, Morgan RD;

XX WPI; 2000-116558/10.

CC Cloning intact genes used to isolate genes for restriction enzymes -  
 Claim 7a; Page 60; 97pp; English.

CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites,  
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
 CC enzymes; toxins that interact with a host, e.g. for synthesis of  
 CC inhibitors or antagonists of the toxin, or in vaccination, or a  
 CC modification methyltransferase. Intact genes can be cloned directly with  
 CC a high probability that the orientation of expression is known in advance  
 CC and low probability of association with extraneous, possibly toxic,  
 CC genes. AA244954-244980 represent the Pseudomonas alcaligenes repeat (PAR)  
 CC elements described in the method of the invention.

XX Sequence 78 BP; 16 A; 25 C; 19 G; 18 T; 0 other;

Query Match 96.9%; Score 50.4; DB 21; Length 78;  
 Best Local Similarity 98.1%; Pred. No. 1e-10;  
 Matches 51; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 52  
 |||||  
 Db 24 tcgttcgctcactgcgggaccggtctaaagcggcccttaaccacagttta 75



RESULT 15  
AAZ44941  
ID AAZ44941 standard; DNA; 79 BP.  
XX  
AC AAZ44941:  
XX  
XX 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes repeat (PAR) element DNA #48.  
XX  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; repeat element; PAR; ss.  
XX  
XX Pseudomonas alcaligenes.  
XX  
XX WO9964632-A1.  
XX  
XX PN 16-DEC-1999.  
XX  
XX PF 11-JUN-1999; 99WO-US13295.  
XX  
XX PR 12-JUN-1998; 98US-0089086.  
XX PR 12-JUN-1998; 98US-0089101.  
XX  
XX (NEW ) NEW ENGLAND BIOIARS INC.  
XX  
XX Raleigh EA, Vaisvila R, Morgan RD;  
XX PI  
XX DR WPI; 2000-116558/10.  
XX  
XX PT Cloning intact genes used to isolate genes for restriction enzymes -  
XX  
XX PS Claim 7a; Page 60; 97pp; English.  
XX  
XX CC This invention describes a novel method for cloning intact,  
XX CC diversity-selected genes (I) from within gene cassettes (GC) which  
XX CC comprises identifying DNA repeats that flank GC, hybridizing  
XX CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
XX CC fragments containing (I), ligating these fragments into a vector and  
XX CC transforming cells with the vector. This method is used to clone a wide  
XX CC variety of prokaryotic genes that provide a selective advantage under  
XX CC particular conditions, particularly those that encode restriction  
XX CC enzymes (used as reagents in molecular biology); adhesins (for use in  
XX CC coating or for targeting molecules or organisms to particular sites,  
XX CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
XX CC enzymes; toxins that interact with a host, e.g. for synthesis of  
XX CC inhibitors or antagonists of the toxin, or in vaccination, or a  
XX CC modification methyltransferase. Intact genes can be cloned directly with  
XX CC a high probability that the orientation of expression is known in advance  
XX CC and low probability of association with extraneous, possibly toxic,  
XX CC genes. AAZ44894-744980 represent the Pseudomonas alcaligenes repeat (PAR)  
XX CC elements described in the method of the invention.  
XX  
SQ Sequence 79 BP; 17 A; 27 C; 18 G; 17 T; 0 other;

Query Match 95.0%; Score 49.4; DB 21; Length 79;  
Best Local Similarity 98.0%; Pred. No. 2.4e-10;  
Matches 50; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 cgccttcgctcactcgggacggcgtaaaagccgcgccttaaccacaaagtta 52  
|||||  
Db 26 cgccttcgctcactcgggacggcgtaaatgcccgccttaaccacaaagtta 76  
|||||

Search completed: August 2, 2001, 16:57:31  
Job time: 261 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 18:40:35 ; Search time 89.09 Seconds  
(without alignments)  
110.497 Million cell updates/sec

Title: US-09-701-626A-79

Perfect score: 52  
Sequence: 1 tcctcgcctcactgcggga.....ggcccttaaccaaacgtta 52

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgnl\_7/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgnl\_7/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgnl\_7/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgnl\_7/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgnl\_7/ptodata/1/ina/PCrUS\_COMB.seq.\*  
6: /cgnl\_7/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	22.6	43.5	1533	3	US-09-122-210-1	Sequence 1, Appli
C 2	22.2	42.7	173	3	US-08-542-051-6	Sequence 6, Appli
C 3	21.2	40.8	7566	2	US-08-232-016-23	Sequence 23, Appli
C 4	21.2	40.8	7639	2	US-08-232-016-22	Sequence 22, Appli
C 5	20.6	39.6	6671	1	US-08-280-443-1	Sequence 1, Appli
C 6	20.6	39.6	6671	1	US-08-457-459-1	Sequence 1, Appli
C 7	20.6	39.6	6671	1	US-08-555-678-1	Sequence 1, Appli
C 8	20.6	39.6	6671	5	PCT-US95-02275-1	Sequence 1, Appli
C 9	20.6	39.6	43280	2	US-08-804-227C-1	Sequence 1, Appli
C 10	20.2	38.8	340	4	US-08-836-075A-57	Sequence 57, Appli
C 11	20.2	38.8	2943	4	US-08-042-747A-7	Sequence 7, Appli
C 12	20.2	38.8	3177	1	US-08-042-747A-4	Sequence 4, Appli
C 13	20	38.5	269	3	US-09-293-823-1	Sequence 1, Appli
C 14	20	38.5	840	5	PCT-US91-08177-12	Sequence 12, Appli
C 15	20	38.5	7323	5	PCT-US91-08177-1	Sequence 1, Appli
C 16	19.8	38.1	340	4	US-08-836-075A-53	Sequence 53, Appli
C 17	19.8	38.1	49377	1	US-08-764-233A-1	Sequence 1, Appli
C 18	19.6	37.7	10095	3	US-08-822-586-45	Sequence 45, Appli
C 19	19.6	37.7	30001	1	US-08-125-468-1	Sequence 1, Appli
C 20	19.6	37.7	30001	2	US-08-474-933-1	Sequence 1, Appli
C 21	19.4	37.3	1001	3	US-08-705-771-6	Sequence 6, Appli
C 22	19.4	37.3	1499	3	US-08-303-861-17	Sequence 17, Appli
C 23	19.4	37.3	1788	3	US-08-303-861-1	Sequence 1, Appli
C 24	19.2	36.9	1893	6	5438126-1	Patent No. 5438126
C 25	19.2	36.9	1958	1	US-08-115-365-1	Sequence 1, Appli
C 26	19.2	36.9	1958	1	US-08-586-897-1	Sequence 1, Appli
C 27	19	36.5	362	4	US-09-060-756-219	Sequence 219, App

```

c 28 19 36.5 397 4 US-09-060-756-285 Sequence 285, App
c 29 19 36.5 851 4 US-08-998-416-331 Sequence 331, App
c 30 19 36.5 1833 3 US-08-704-966-3 Sequence 3, Appli
c 31 19 36.5 1833 3 US-08-705-438-3 Sequence 1, Appli
c 32 19 36.5 3758 3 US-08-323-477-1 Sequence 1, Appli
c 33 19 36.5 5996 3 US-09-028-934-33 Sequence 33, Appli
c 34 18.8 36.2 86 1 US-08-447-169A-214 Sequence 214, App
c 35 18.8 36.2 1392 4 US-09-163-444-1 Sequence 1, Appli
c 36 18.8 36.2 1608 1 US-07-621-670-2 Sequence 2, Appli
c 37 18.8 36.2 1735 4 US-09-163-444-3 Sequence 3, Appli
c 38 18.8 36.2 2145 1 US-08-849-212-5 Sequence 5, Appli
c 39 18.8 36.2 2703 2 US-08-288-508C-1 Sequence 1, Appli
c 40 18.8 36.2 3252 2 US-08-809-740A-1 Sequence 1, Appli
c 41 18.8 36.2 3252 2 US-08-809-740A-4 Sequence 4, Appli
c 42 18.8 36.2 8438 1 US-07-945-283-1 Sequence 1, Appli
c 43 18.8 36.2 15202 3 US-08-922-635-21 Sequence 21, Appli
c 44 18.6 35.8 340 3 US-08-441-971-9 Sequence 9, Appli
c 45 18.6 35.8 340 4 US-08-442-144A-10 Sequence 10, Appli
```

#### ALIGNMENTS

```

RESULT 1
US-09-122-210-1/c
; Sequence 1, Application US/09122210
; Patent No. 6140083
; GENERAL INFORMATION:
; APPLICANT: Haake, David A.
; TITLE OF INVENTION: Leptospiral Outer Membrane Protein
; FILE REFERENCE: 09638/005001
; CURRENT APPLICATION NUMBER: US/09/122,210
; CURRENT FILING DATE: 1998-07-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1533
; TYPE: DNA
; ORGANISM: Leptospira sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(1463)
US-09-122-210-1
```

```

Query Match 43.5%; Score 22.6; DB 3; Length 1533;
Best Local Similarity 68.9%; Pred. No. 3.3;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

```

QY 7 cgctcactgcgggacgcggtaaagccgccttaaccacacgtt 51
      ||||| |||| |||| |||| |||| |||| |||| ||||
DB 1362 CGCTTAGACGGCAACCGGCTAAGCCGGCGCTCTCCAGAGGTT 1318
```

```

RESULT 2
US-08-542-051-6/c
; Sequence 6, Application US/08542051F
; Patent No. 6004782
; GENERAL INFORMATION:
; APPLICANT: Daniel, Henry
; APPLICANT: McPherson, David T.
; APPLICANT: Urry, Dan W.
; APPLICANT: Xu, Jie
; TITLE OF INVENTION: Hyperexpression of Bioelastic Polypeptides
; FILE REFERENCE: BERL-018/01US
; CURRENT APPLICATION NUMBER: US/08/542,051F
; CURRENT FILING DATE: 1995-10-13
; EARLIER APPLICATION NUMBER: 08/423,642
; EARLIER FILING DATE: 1995-04-14
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 173
```



```

RESULT      5
US-08-443-1/c
; Sequence 1, Application US/08280443
; Patent No. 5643778
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,443
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49AUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6671 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 155..3832
; US-08-280-443-1

Query Match          39.6%; Score 20.6; DB 1; Length 6671
Best Local Similarity 67.4%; Pred.No.24;
Matches 25; Conservative 0; Mismatches 14; Indels

QY    7   cgctcactgcgggacgcgttaaacgccgccttaccacaacg 49
       || || |||| |||| | |||| |||| |||| |||| ||||
Db    3090 CGGTCGTGCAGGACTTGTCAAAGAGGGCGCCATCTCCACG 3048

RESULT      6
US-08-457-459-1/c
; Sequence 1, Application US/08457459
; Patent No. 5677428
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA

```

```

; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
;
; TITLE OF INVENTION: Thereof
;
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA

```

```

? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/280,443
? FILING DATE: 25-JUL-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/457,459
? FILING DATE: 01-JUN-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Bak, Mary E.
? REGISTRATION NUMBER: 31,215
? REFERENCE/DOCKET NUMBER: WST49DUSA
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 215-540-9206
? TELEFAX: 215-540-5818
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 6671 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 155..3832
? US-08-555-678-1

Query Match 39.6%; Score 20.6; DB 1; Length 6671;
Best Local Similarity 67.4%; Pred. No. 24;
Matches 29; Conservative 0; Mismatches 14; Indels 0

Qy 7 cgctcactgcgggacgcgctaagacggcccttaacaaacg 49
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3090 CGGTGCGTGCAGGACTTGTCAAAGAGGCGCCATCTCCACAC 3048

```

```

1  RESULT      8
2  PCT-US95-02275-1/C
3  : Sequence 1, Application PC/TUS9502275
4  : GENERAL INFORMATION:
5  : APPLICANT: Wiscar Institute of Anatomy & Biology
6  : TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
7  : TITLE OF INVENTION: Thereof
8  : NUMBER OF SEQUENCES: 39
9  : CORRESPONDENCE ADDRESS:
10 : ADDRESSEE: Howson and Howson
11 : STREET: Spring House Corporate Cntr, P.O. Box 457
12 : CITY: Spring House
13 : STATE: Pennsylvania
14 : COUNTRY: USA
15 : ZIP: 19477
16 : COMPUTER READABLE FORM:
17 : MEDIUM TYPE: Floppy disk
18 : COMPUTER: IBM PC compatible
19 : OPERATING SYSTEM: PC-DOS/MS-DOS
20 : SOFTWARE: Patent In Release #1.0, Version #1.30
21 : CURRENT APPLICATION DATA:
22 : APPLICATION NUMBER: PCT/US95/02275
23 : FILING DATE:
24 : CLASSIFICATION:
25 : PRIOR APPLICATION DATA:
26 : APPLICATION NUMBER: US 08/280,443
27 : FILING DATE: 25-JUL-1994
28 : PRIOR APPLICATION DATA:
29 : APPLICATION NUMBER: US 08/197,794
30 : FILING DATE: 17-FEB-1994
31 : ATTORNEY/AGENT INFORMATION:
32 : NAME: Bak, Mary E.
33 : REGISTRATION NUMBER: 31,215
34 : REFERENCE/DOCKET NUMBER: WST49BPCT
35 : TELECOMMUNICATION INFORMATION:
36 : TELEPHONE: 215-540-9206
37 : TELEFAX: 215-540-5818
38 : INFORMATION FOR SI ID NO: 1:

```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 6671 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 155..3832
: PCT-US95-02275-1

Query Match          39.6%; Score 20.6; DB 5; Length 6671;
Best local Similarity 67.4%; Pred. No. 24;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 7 cgtcactcgggaccgctaaagccgccttaacaaacy 49
DB 3090 CGTCTCGTCGAGGACTGTCTAAAGAGGCGCCCATCTCCACAG 3048

RESULT 9
US-08-804-227C-1
: Sequence 1, Application US/08804227C
: Patent No. 5876991
: GENERAL INFORMATION:
: APPLICANT: Delhoff, Bradley S.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rostock, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) Text only
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELEPHONE: 317-276-2459
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 43280 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 816..14234
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 14351..19945
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 20010..31199
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 31232..36067
```

```

: FEATURE:
: NAME/KEY: CDS
: LOCATION: 36249..41774
: US-08-804-227C-1

Query Match          39.6%; Score 20.6; DB 2; Length 43280;
Best local Similarity 74.3%; Pred. No. 13;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 qcttcactcactcgggaccgctaaagccgcctc 37
DB 24347 GCTCGCTCTCTCCGGGCGCCGCGAGGAGGCC 24381

RESULT 10
US-08-836-075A-57
: Sequence 57, Application US/08836075A
: Patent No. 6180768
: GENERAL INFORMATION:
: APPLICANT: MAERTENS, GEERT
: APPLICANT: STUYVER, LIEVEN
: TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
: TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
: TITLE OF INVENTION: AGENTS
: NUMBER OF SEQUENCES: 207
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ARNOLD, WHITE & DURKEE
: STREET: P.O. BOX 4433
: CITY: HOUSTON
: STATE: TEXAS
: COUNTRY: USA
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Microsoft Word 6.0 / ASCII text output
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/836,075A
: FILING DATE: 21 Apr 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/EP95/04155
: FILING DATE: 23 Oct 1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EP 94870166.9
: FILING DATE: 21 Oct 1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EP 95870076.7
: FILING DATE: 28 Jun 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: KAMMERER, PATRICIA A.
: REGISTRATION NUMBER: 29,775
: REFERENCE/DOCKET NUMBER: INNS:004
: INFORMATION FOR SEQ ID NO: 57:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 340 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US-08-836-075A-57

Query Match          38.8%; Score 20.2; DB 4; Length 340;
Best local Similarity 68.3%; Pred. No. 22;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 6 tcgtcactcgggaccgctaaagccgccttaaccaa 46
DB 89 TCGCTCACCAGCGGCTGTATATCGGGGTCCCTAACCAA 129
```

```
RESULT 11
US-08-042-747A-7
; Sequence 7, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scinicariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East Pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042.747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2943 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 87..2744
; US-08-042-747A-7

Query Match 38.8%; Score 20.2; DB 1; Length 2943;
Best Local Similarity 75.8%; Pred. No. 31;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Caps 0;

QY 7 cgtcactgcgggaccggtacaaagccgccct 39
Db 1811 CGCGCGCCCGGGAGCTGCTACACGCCGCCCT 1843

RESULT 12
US-08-042-747A-4
; Sequence 4, Application US/08042747A
; Patent No. 5487969
; GENERAL INFORMATION:
; APPLICANT: Eberle, Richard
; APPLICANT: Black, Darla
; APPLICANT: Scinicariello, Franco
; APPLICANT: Hilliard, Julia K.
; TITLE OF INVENTION: Cloning and Amplification of Monkey B
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Cox & Smith Incorporated
; STREET: 112 East Pecan Street, Suite 2000
; CITY: San Antonio
; STATE: Texas
; COUNTRY: USA
; ZIP: 78205
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/042.747A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haymond, W. Bradley
; REGISTRATION NUMBER: 35186
; REFERENCE/DOCKET NUMBER: S-0072.179
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 210-554-5500
; TELEFAX: 210-226-8395
; TELEX: 767609
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3177 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 269..2941
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..249
; US-08-042-747A-4

Query Match 38.8%; Score 20.2; DB 1; Length 3177;
Best Local Similarity 75.8%; Pred. No. 31;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Caps 0;

QY 7 cgtcactgcgggaccggtacaaagccgccct 39
Db 1999 CGCGCGCCCGGGAGCTGCTACACGCCGCCCT 2031

RESULT 13
US-09-293-823-1/c
; Sequence 1, Application US/09293823
; Patent No. 6126955
; GENERAL INFORMATION:
; APPLICANT: William D. Huse
; TITLE OF INVENTION: A NEW METHOD FOR TAPPING THE
; FILE REFERENCE: 62710024US06
; CURRENT APPLICATION NUMBER: US/09/293,823
; CURRENT FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Synthetic DNA
; US-09-293-823-1

Query Match 38.5%; Score 20; DB 3; Length 269;
Best Local Similarity 72.2%; Pred. No. 25;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Caps 0;

QY 8 gctcactgcgggaccggtacaaagccgcccttaac 43
```





Best Local Similarity 65.9%; Pred. No. 42;  
Matches 29; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
Qy 5 ttcgctcactgaggaccggctaaagccggcccttaaccaaac 48  
|| | | | | | | | | | | | | | | | | | | | |  
Db 3743 TTACATTAGTGTGGAAATGCTAAAGCCAGCCACATCCAAC 3786

Search completed: August 2, 2001, 18:40:43  
Job time: 6333 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 17:44:12 Search time 2788.81 Seconds  
(without alignments)  
275.528 Million cell updates/sec

Title: US-09-701-626A-79  
Perfect score: 52  
Sequence: 1 tcgcttcgctcactgcggga.....ggcccttaaccacagtta 52

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 14155048 seqs, 7388405095 residues  
Total number of hits satisfying chosen parameters: 28310096

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main:\*  
1: /cgnl\_7/ptodata/1/pna/PCTUS\_COMB.seq:\*  
2: /cgnl\_7/ptodata/1/pna/US06\_COMB.seq:\*  
3: /cgnl\_7/ptodata/1/pna/US07\_COMB.seq:\*  
4: /cgnl\_7/ptodata/1/pna/US080\_COMB.seq:\*  
5: /cgnl\_7/ptodata/1/pna/US081\_COMB.seq:\*  
6: /cgnl\_7/ptodata/1/pna/US082\_COMB.seq:\*  
7: /cgnl\_7/ptodata/1/pna/US083\_COMB.seq:\*  
8: /cgnl\_7/ptodata/1/pna/US084\_COMB.seq:\*  
9: /cgnl\_7/ptodata/1/pna/US085\_COMB.seq:\*  
10: /cgnl\_7/ptodata/1/pna/US086\_COMB.seq:\*  
11: /cgnl\_7/ptodata/1/pna/US087\_COMB.seq:\*  
12: /cgnl\_7/ptodata/1/pna/US088\_COMB.seq:\*  
13: /cgnl\_7/ptodata/1/pna/US089\_COMB.seq:\*  
14: /cgnl\_7/ptodata/1/pna/US090\_COMB.seq:\*  
15: /cgnl\_7/ptodata/1/pna/US091\_COMB.seq:\*  
16: /cgnl\_7/ptodata/1/pna/US092\_COMB.seq:\*  
17: /cgnl\_7/ptodata/1/pna/US093\_COMB.seq:\*  
18: /cgnl\_7/ptodata/1/pna/US094\_COMB.seq:\*  
19: /cgnl\_7/ptodata/1/pna/US095A\_COMB.seq:\*  
20: /cgnl\_7/ptodata/1/pna/US095B\_COMB.seq:\*  
21: /cgnl\_7/ptodata/1/pna/US095C\_COMB.seq:\*  
22: /cgnl\_7/ptodata/1/pna/US095D\_COMB.seq:\*  
23: /cgnl\_7/ptodata/1/pna/US096A\_COMB.seq:\*  
24: /cgnl\_7/ptodata/1/pna/US096B\_COMB.seq:\*  
25: /cgnl\_7/ptodata/1/pna/US096C\_COMB.seq:\*  
26: /cgnl\_7/ptodata/1/pna/US096D\_COMB.seq:\*  
27: /cgnl\_7/ptodata/1/pna/US096E\_COMB.seq:\*  
28: /cgnl\_7/ptodata/1/pna/US097A\_COMB.seq:\*  
29: /cgnl\_7/ptodata/1/pna/US097B\_COMB.seq:\*  
30: /cgnl\_7/ptodata/1/pna/US097C\_COMB.seq:\*  
31: /cgnl\_7/ptodata/1/pna/US098\_COMB.seq:\*  
32: /cgnl\_7/ptodata/1/pna/US6000\_COMB.seq:\*  
33: /cgnl\_7/ptodata/1/pna/US6001\_COMB.seq:\*  
34: /cgnl\_7/ptodata/1/pna/US6002\_COMB.seq:\*  
35: /cgnl\_7/ptodata/1/pna/US6003\_COMB.seq:\*  
36: /cgnl\_7/ptodata/1/pna/US6004\_COMB.seq:\*  
37: /cgnl\_7/ptodata/1/pna/US6005\_COMB.seq:\*  
38: /cgnl\_7/ptodata/1/pna/US6006\_COMB.seq:\*  
39: /cgnl\_7/ptodata/1/pna/US6007\_COMB.seq:\*  
40: /cgnl\_7/ptodata/1/pna/US6008\_COMB.seq:\*  
41: /cgnl\_7/ptodata/1/pna/US6009\_COMB.seq:\*  
42: /cgnl\_7/ptodata/1/pna/US6010\_COMB.seq:\*  
43: /cgnl\_7/ptodata/1/pna/US6011\_COMB.seq:\*

44: /cgnl\_7/ptodata/1/pna/US6012\_COMB.seq:\*  
45: /cgnl\_7/ptodata/1/pna/US6013\_COMB.seq:\*  
46: /cgnl\_7/ptodata/1/pna/US6014\_COMB.seq:\*  
47: /cgnl\_7/ptodata/1/pna/US6015\_COMB.seq:\*  
48: /cgnl\_7/ptodata/1/pna/US6016\_COMB.seq:\*  
49: /cgnl\_7/ptodata/1/pna/US6017\_COMB.seq:\*  
50: /cgnl\_7/ptodata/1/pna/US6018\_COMB.seq:\*  
51: /cgnl\_7/ptodata/1/pna/US6019\_COMB.seq:\*  
52: /cgnl\_7/ptodata/1/pna/US6020\_COMB.seq:\*  
53: /cgnl\_7/ptodata/1/pna/US6021\_COMB.seq:\*  
54: /cgnl\_7/ptodata/1/pna/US6022\_COMB.seq:\*  
55: /cgnl\_7/ptodata/1/pna/US6023\_COMB.seq:\*  
56: /cgnl\_7/ptodata/1/pna/US6024\_COMB.seq:\*  
57: /cgnl\_7/ptodata/1/pna/US6025\_COMB.seq:\*  
58: /cgnl\_7/ptodata/1/pna/US6026\_COMB.seq:\*  
59: /cgnl\_7/ptodata/1/pna/US6027\_COMB.seq:\*  
60: /cgnl\_7/ptodata/1/pna/US6028\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	52	100.0	52	1	PCT-US99-13295-79 Sequence 79, Appl
2	52	100.0	52	28	US-09-701-626A-79 Sequence 23, Appl
3	52	100.0	78	1	PCT-US99-13295-23 Sequence 38, Appl
4	52	100.0	78	1	PCT-US99-13295-38 Sequence 47, Appl
5	52	100.0	78	1	PCT-US99-13295-47 Sequence 49, Appl
6	52	100.0	78	1	PCT-US99-13295-49 Sequence 51, Appl
7	52	100.0	78	1	PCT-US99-13295-51 Sequence 59, Appl
8	52	100.0	78	1	PCT-US99-13295-59 Sequence 23, Appl
9	52	100.0	78	28	US-09-701-626A-23 Sequence 38, Appl
10	52	100.0	78	28	US-09-701-626A-38 Sequence 47, Appl
11	52	100.0	78	28	US-09-701-626A-47 Sequence 49, Appl
12	52	100.0	78	28	US-09-701-626A-49 Sequence 51, Appl
13	52	100.0	78	28	US-09-701-626A-51 Sequence 59, Appl
14	52	100.0	78	28	US-09-701-626A-59 Sequence 3, Appl
15	52	100.0	7300	28	US-09-701-626A-2 Sequence 3, Appl
16	52	100.0	7300	28	US-09-701-626A-3 Sequence 4, Appl
17	52	100.0	12956	28	US-09-701-626A-4 Sequence 1, Appl
18	52	100.0	14143	1	PCT-US99-13295-1 Sequence 2, Appl
19	52	100.0	14143	1	PCT-US99-13295-2 Sequence 1, Appl
20	52	100.0	14143	28	US-09-701-626A-1 Sequence 4, Appl
21	52	100.0	42143	1	PCT-US99-13295-67 Sequence 67, Appl
22	52	100.0	42143	1	PCT-US99-13295-67 Sequence 52, Appl
23	50.4	96.9	78	1	US-09-701-626A-52 Sequence 11, Appl
24	50.4	96.9	78	28	US-09-701-626A-67 Sequence 19, Appl
25	49.4	95.0	79	1	PCT-US99-13295-52 Sequence 22, Appl
26	49.4	95.0	79	28	US-09-701-626A-52 Sequence 24, Appl
27	40	76.9	77	1	PCT-US99-13295-11 Sequence 55, Appl
28	40	76.9	77	1	PCT-US99-13295-19 Sequence 63, Appl
29	40	76.9	77	1	PCT-US99-13295-22 Sequence 65, Appl
30	40	76.9	77	1	PCT-US99-13295-24 Sequence 77, Appl
31	40	76.9	77	1	PCT-US99-13295-41 Sequence 78, Appl
32	40	76.9	77	1	PCT-US99-13295-55 Sequence 119, App
33	40	76.9	77	1	PCT-US99-13295-63 Sequence 11, Appl
34	40	76.9	77	1	PCT-US99-13295-65 Sequence 19, Appl
35	40	76.9	77	1	PCT-US99-13295-68 Sequence 22, Appl
36	40	76.9	77	1	PCT-US99-13295-77 Sequence 41, Appl
37	40	76.9	77	1	PCT-US99-13295-78 Sequence 55, Appl
38	40	76.9	77	1	PCT-US99-13295-78 Sequence 63, Appl
39	40	76.9	77	28	US-09-701-626A-11 Sequence 65, Appl
40	40	76.9	77	28	US-09-701-626A-19 Sequence 77, Appl
41	40	76.9	77	28	US-09-701-626A-22 Sequence 11, Appl
42	40	76.9	77	28	US-09-701-626A-24 Sequence 19, Appl
43	40	76.9	77	28	US-09-701-626A-41 Sequence 22, Appl
44	40	76.9	77	28	US-09-701-626A-41 Sequence 24, Appl
45	40	76.9	77	28	US-09-701-626A-55 Sequence 41, Appl
					Sequence 55, Appl
					Sequence 119, App

## ALIGNMENTS

RESULT 1  
PCT-US99-13295-79  
; Sequence 79, Application PC/TUS9913295  
; GENERAL INFORMATION:  
; APPLICANT: New England Biolabs, Inc.  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: Gene Discovery Method  
; CURRENT APPLICATION NUMBER: PCT/US99/13295  
; CURRENT FILING DATE: 1999-06-11  
; EARLIER APPLICATION NUMBER: 60/089,086  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER APPLICATION NUMBER: 60/089,101  
; EARLIER FILING DATE: 1998-06-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 79  
; LENGTH: 52  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Synthetic  
; OTHER INFORMATION: Oligonucleotide based on Pseudomonas Alcaligenes  
; OTHER INFORMATION: NEB#585 (ATCC 55044)  
PCT-US99-13295-79

Query Match 100.0%; Score 52; DB 1; Length 52;  
Best Local Similarity 100.0%; Pred. No. 8.2e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgcttcgctcactgcgggaccggtctaaagccgccccttaacaaacgtta 52  
|||||  
Db 1 tcgcttcgctcactgcgggaccggtctaaagccgccccttaacaaacgtta 52

RESULT 2  
US-09-701-626A-79  
; Sequence 79, Application US/09701626A  
; GENERAL INFORMATION:  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: NEB-165PUS  
; CURRENT APPLICATION NUMBER: US/09/701,626A  
; CURRENT FILING DATE: 2001-04-16  
; PRIOR APPLICATION NUMBER: 60/089,086  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089,101  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: PCT/US99/13295  
; PRIOR FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 79  
; LENGTH: 52  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Synthetic  
; OTHER INFORMATION: Oligonucleotide based on Pseudomonas Alcaligenes  
; OTHER INFORMATION: NEB#585 (ATCC 55044)  
US-09-701-626A-79

Query Match 100.0%; Score 52; DB 28; Length 52;

Best Local Similarity 100.0%; Pred. No. 8.2e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgcttcgctcactgcgggaccggtctaaagccgccccttaacaaacgtta 52  
|||||  
Db 1 tcgcttcgctcactgcgggaccggtctaaagccgccccttaacaaacgtta 52

RESULT 3  
PCT-US99-13295-23  
; Sequence 23, Application PC/TUS9913295  
; GENERAL INFORMATION:  
; APPLICANT: New England Biolabs, Inc.  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: Gene Discovery Method  
; CURRENT APPLICATION NUMBER: PCT/US99/13295  
; CURRENT FILING DATE: 1999-06-11  
; EARLIER APPLICATION NUMBER: 60/089,086  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER APPLICATION NUMBER: 60/089,101  
; EARLIER FILING DATE: 1998-06-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 78  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)  
PCT-US99-13295-23

Query Match 100.0%; Score 52; DB 1; Length 78;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tcgcttcgctcactgcgggaccggtctaaagccgccccttaacaaacgtta 52  
|||||  
Db 24 tcgcttcgctcactgcgggaccggtctaaagccgccccttaacaaacgtta 75

RESULT 4  
PCT-US99-13295-38  
; Sequence 38, Application PC/TUS9913295  
; GENERAL INFORMATION:  
; APPLICANT: New England Biolabs, Inc.  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: Gene Discovery Method  
; CURRENT APPLICATION NUMBER: PCT/US99/13295  
; CURRENT FILING DATE: 1999-06-11  
; EARLIER APPLICATION NUMBER: 60/089,086  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER APPLICATION NUMBER: 60/089,101  
; EARLIER FILING DATE: 1998-06-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 38  
; LENGTH: 78  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)  
PCT-US99-13295-38

Query Match 100.0%; Score 52; DB 1; Length 78;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 52  
|||||  
Db 24 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 75  
|||||

## RESULT 5

PCT-US99-13295-47  
; Sequence 47, Application PC/TUS9913295  
; GENERAL INFORMATION:  
; APPLICANT: New England Biolabs, Inc.  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: Gene Discovery Method  
; CURRENT APPLICATION NUMBER: PCT/US99/13295  
; EARLIER FILING DATE: 1999-06-11  
; EARLIER APPLICATION NUMBER: 60/089,086  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER APPLICATION NUMBER: 60/089,101  
; EARLIER FILING DATE: 1998-06-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 47  
; LENGTH: 78  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)  
PCT-US99-13295-47

Query Match 100.0%; Score 52; DB 1; Length 78;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 52  
|||||  
Db 24 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 75  
|||||

## RESULT 6

PCT-US99-13295-49  
; Sequence 49, Application PC/TUS9913295  
; GENERAL INFORMATION:  
; APPLICANT: New England Biolabs, Inc.  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: Gene Discovery Method  
; CURRENT APPLICATION NUMBER: PCT/US99/13295  
; EARLIER FILING DATE: 1999-06-11  
; EARLIER APPLICATION NUMBER: 60/089,086  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER APPLICATION NUMBER: 60/089,101  
; EARLIER FILING DATE: 1998-06-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 49  
; LENGTH: 78  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)  
PCT-US99-13295-49

Query Match 100.0%; Score 52; DB 1; Length 78;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 52  
|||||  
Db 24 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 75  
|||||

## RESULT 7

PCT-US99-13295-51  
; Sequence 51, Application PC/TUS9913295  
; GENERAL INFORMATION:  
; APPLICANT: New England Biolabs, Inc.  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: Gene Discovery Method  
; CURRENT APPLICATION NUMBER: PCT/US99/13295  
; EARLIER FILING DATE: 1999-06-11  
; EARLIER APPLICATION NUMBER: 60/089,086  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER APPLICATION NUMBER: 60/089,101  
; EARLIER FILING DATE: 1998-06-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 51  
; LENGTH: 78  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)  
PCT-US99-13295-51

Query Match 100.0%; Score 52; DB 1; Length 78;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 52  
|||||  
Db 24 tcgcttcgctcactcgggaccggtctaaagccggcccttaacccaaacgtta 75  
|||||

## RESULT 8

PCT-US99-13295-59  
; Sequence 59, Application PC/TUS9913295  
; GENERAL INFORMATION:  
; APPLICANT: New England Biolabs, Inc.  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: Gene Discovery Method  
; CURRENT APPLICATION NUMBER: PCT/US99/13295  
; EARLIER FILING DATE: 1999-06-11  
; EARLIER APPLICATION NUMBER: 60/089,086  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER APPLICATION NUMBER: 60/089,101  
; EARLIER FILING DATE: 1998-06-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 59  
; LENGTH: 78  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)  
PCT-US99-13295-59

```
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-38

Query Match          100.0%; Score 52; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 52
   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 24 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 75

RESULT 9
US-09-701-626A-23
; Sequence 23, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Valsvilla, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-23

Query Match          100.0%; Score 52; DB 28; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 52
   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 24 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 75

RESULT 10
US-09-701-626A-38
; Sequence 38, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Valsvilla, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-47

Query Match          100.0%; Score 52; DB 28; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 52
   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 24 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 75

RESULT 11
US-09-701-626A-47
; Sequence 47, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Valsvilla, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-47

Query Match          100.0%; Score 52; DB 28; Length 78;
Best Local Similarity 100.0%; Pred. No. 8.5e-10;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 52
   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 24 tcgcttcgctcaactgcgggaccggtctaaagccggcccttaacccaacgtta 75

RESULT 12
US-09-701-626A-49
; Sequence 49, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Valsvilla, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
```

; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:

; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)  
US-09-701-626A-49

Query Match 100.0%; Score 52; DB 28; Length 78;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 52  
|||||  
Db 24 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 75

## RESULT 13

US-09-701-626A-51

; Sequence 51, Application US/09701626A

; GENERAL INFORMATION:

; APPLICANT: Vaisvila, Romualdas

; APPLICANT: Morgan, Richard D.

; APPLICANT: Raleigh, Elisabeth

; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method

; FILE REFERENCE: NEB-165PUS

; CURRENT APPLICATION NUMBER: US/09/701,626A

; CURRENT FILING DATE: 2001-04-16

; PRIOR FILING DATE: 1998-06-12

; PRIOR APPLICATION NUMBER: 60/089,086

; PRIOR FILING DATE: 1998-06-12

; PRIOR APPLICATION NUMBER: 60/089,101

; PRIOR FILING DATE: 1998-06-12

; PRIOR APPLICATION NUMBER: PCT/US99/13295

; PRIOR FILING DATE: 1999-06-11

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 51

; LENGTH: 78

; TYPE: DNA

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of

; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)

US-09-701-626A-51

Query Match 100.0%; Score 52; DB 28; Length 78;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 52  
|||||  
Db 24 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 75

## RESULT 14

US-09-701-626A-59

; Sequence 59, Application US/09701626A

; GENERAL INFORMATION:

; APPLICANT: Vaisvila, Romualdas

; APPLICANT: Morgan, Richard D.

; APPLICANT: Raleigh, Elisabeth

; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method

; FILE REFERENCE: NEB-165PUS

; CURRENT APPLICATION NUMBER: US/09/701,626A

; CURRENT FILING DATE: 2001-04-16

; PRIOR FILING DATE: 1998-06-12

; PRIOR APPLICATION NUMBER: 60/089,086

; PRIOR FILING DATE: 1998-06-12

; PRIOR APPLICATION NUMBER: 60/089,101

; PRIOR FILING DATE: 1998-06-12

; PRIOR APPLICATION NUMBER: PCT/US99/13295

; PRIOR FILING DATE: 1999-06-11

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 59

; LENGTH: 78

; TYPE: DNA

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of

; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)

US-09-701-626A-59

Query Match 100.0%; Score 52; DB 28; Length 78;  
Best Local Similarity 100.0%; Pred. No. 8.5e-10;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 52  
|||||  
Db 24 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 75

## RESULT 15

PCT-US99-13295-3

; Sequence 3, Application PC/TUS9913295

; GENERAL INFORMATION:

; APPLICANT: New England Biolabs, Inc.

; APPLICANT: Vaisvila, Romualdas

; APPLICANT: Morgan, Richard D.

; APPLICANT: Raleigh, Elisabeth

; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method

; FILE REFERENCE: Gene Discovery Method

; CURRENT APPLICATION NUMBER: PCT/US99/13295

; CURRENT FILING DATE: 1999-06-11

; EARLIER APPLICATION NUMBER: 60/089,086

; EARLIER FILING DATE: 1998-06-12

; EARLIER APPLICATION NUMBER: 60/089,101

; EARLIER FILING DATE: 1998-06-12

; NUMBER OF SEQ ID NOS: 130

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 7300

; TYPE: DNA

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of

; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)

PCT-US99-13295-3

Query Match 100.0%; Score 52; DB 1; Length 7300;  
Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 52  
|||||  
Db 3300 tcgcttcgctcactgcgggaccggtctaaagccggcccttaacccaaacgtta 3351

Search completed: August 2, 2001, 17:44:12  
Job time: 3062 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:54:36 ; Search time 49.77 Seconds  
(without alignments)  
730.055 Million cell updates/sec

Title: US-09-701-626A-79

Perfect score: 52  
Sequence: 1 tcgttcgctcactgcggga.....ggcccttaaccaaacgtta 52

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 327855 seqs, 34937361 residues

Total number of hits satisfying chosen parameters: 655710

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents NA.New.\*

- 1: /cgnl\_7/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*
- 2: /cgnl\_7/ptodata/2/pna/US06\_NEW\_COMB.seq.\*
- 3: /cgnl\_7/ptodata/2/pna/US07\_NEW\_COMB.seq.\*
- 4: /cgnl\_7/ptodata/2/pna/US08\_NEW\_COMB.seq.\*
- 5: /cgnl\_7/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 6: /cgnl\_7/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.4	45.0	422	5	US-09-804-730-10706
2	22.4	43.1	388	5	US-09-904-703-1102
3	22.4	43.1	531	5	US-09-904-703-14754
4	22.4	43.1	1409	5	US-09-758-474-201
5	22.2	42.7	837	5	US-09-803-110-6363
6	22.2	42.7	1395	5	US-09-803-110-4177
7	22.2	42.7	3157	5	US-09-764-902-493
8	22.2	42.7	8416	6	US-60-278-232-8896
9	22.2	42.7	12051	5	US-09-764-902-2125
10	22.2	42.7	18416	5	US-09-764-902-2124
11	22.2	42.7	260744	5	US-09-803-110-211
12	22.2	42.7	506113	5	US-09-803-110-218
13	21.6	41.5	1029	5	US-09-803-110-3210
14	21.6	41.5	1145	5	US-09-758-457-139
15	21.6	41.5	4202	6	US-60-278-232-2701
16	21.6	41.5	78844	5	US-09-803-736-726
17	21.6	41.5	184668	5	US-09-803-110-207
18	21.2	40.8	3207	5	US-09-803-110-3990
19	21.2	40.8	244769	5	US-09-803-110-210
20	21.0	40.4	4840	6	US-60-278-258-11152
21	20.8	40.0	315	5	US-09-757-032-605
22	20.8	40.0	86014	5	US-09-803-736-1009
23	20.6	39.6	389	5	US-09-804-730-13213
24	20.6	39.6	393	5	US-09-803-110-6830
25	20.6	39.6	398	5	US-09-904-703-6327
26	20.6	39.6	414	5	US-09-764-902-222
27	20.6	39.6	2641	5	US-09-875-195-314

28	20.6	39.6	2641	5	US-09-758-471-1302	Sequence 1302, Ap
c 29	20.6	39.6	5271	6	US-60-278-258-4682	Sequence 4682, Ap
c 30	20.6	39.6	6635	6	US-60-278-232-4263	Sequence 4263, Ap
31	20.6	39.6	7010	5	US-09-764-891-6031	Sequence 6031, Ap
c 32	20.6	39.6	19616	5	US-09-764-877-3220	Sequence 3220, Ap
33	20.6	39.6	93672	5	US-09-803-736-1459	Sequence 1459, Ap
c 34	20.6	39.6	107931	5	US-09-803-736-752	Sequence 752, App
35	20.4	39.2	288	5	US-09-803-110-6088	Sequence 6088, Ap
c 36	20.4	39.2	397803	5	US-09-803-110-217	Sequence 217, App
c 37	20.2	38.8	482	5	US-09-823-301-2859	Sequence 2859, Ap
38	20.2	38.8	517	5	US-09-758-460-28	Sequence 28, Appl
c 39	20.2	38.8	254289	5	US-09-803-110-212	Sequence 212, App
c 40	20	38.5	108582	5	US-09-803-736-1236	Sequence 1236, Ap
41	20	38.5	110514	5	US-09-803-736-132	Sequence 132, App
c 42	20	38.5	113172	5	US-09-803-736-113	Sequence 113, App
c 43	20	38.5	118737	5	US-09-803-736-553	Sequence 553, App
44	19.8	38.1	605	6	US-60-255-619-12304	Sequence 12304, A
45	19.8	38.1	755	5	US-09-758-472-1387	Sequence 1387, Ap

#### ALIGNMENTS

RESULT 1  
US-09-804-730-10706  
; Sequence 10706, Application US/09804730  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Scott E.  
; APPLICANT: Hammond-Kosack, Kim  
; APPLICANT: Masucci, James D.  
; APPLICANT: Urban, Martin  
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
; FILE REFERENCE: 38-21(51838)B  
; CURRENT APPLICATION NUMBER: US/09/804,730  
; PRIOR FILING DATE: 2001-03-13  
; PRIOR APPLICATION NUMBER: US 60/189,657  
; PRIOR FILING DATE: 2000-03-15  
; NUMBER OF SEQ ID NOS: 22828  
; SEQ ID NO 10706  
; LENGTH: 422  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(422)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: LIB3352-021-PI-K1-G5  
US-09-804-730-10706

Query Match 45.0%; Score 23.4; DB 5; Length 422;  
Best Local Similarity 67.3%; Pred. No. 3.9;  
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;  
QY 3 gcttcgctcactgcggagcggctaaagccgcgccttaaccacagtt 51  
Db 74 gcttccttcgactgcgggcccgcctcccaagcagctccttccacgaaatctt 122

RESULT 2  
US-09-904-703-1102/c  
; Sequence 1102, Application US/09904703  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-758CON1  
; CURRENT APPLICATION NUMBER: US/09/904,703  
; CURRENT FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: 09/210,298  
; PRIOR FILING DATE: 1998-12-09  
; NUMBER OF SEQ ID NOS: 17812



```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1102
; LENGTH: 388
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-904-703-1102

Query Match          43.1%; Score 22.4; DB 5; Length 388;
Best Local Similarity 66.7%; Pred. NO. 9.5;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 cgcctgcctcactgcgggaccggtctaaagccggcccttaacccaacg 49
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 137 CCCTTCACACACTGATGGACACGCTACACAGGAGCGATACAAAGG 90

RESULT 3
US-09-904-703-14754/c
; Sequence 14754, Application US/09904703
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-758CON1
; CURRENT APPLICATION NUMBER: US/09/904,703
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 09/210,298
; PRIOR FILING DATE: 1998-12-09
; NUMBER OF SEQ ID NOS: 17812
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14754
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(531)
; OTHER INFORMATION: n = A,T,C or G
US-09-904-703-14754

Query Match          43.1%; Score 22.4; DB 5; Length 531;
Best Local Similarity 66.7%; Pred. NO. 10;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 cgcctgcctcactgcgggaccggtctaaagccggcccttaacccaacg 49
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 216 CCCTTCACACACTGATGGACACGCTACACAGGAGCGATACAAAGG 169

RESULT 4
US-09-758-474-201/c
; Sequence 201, Application US/09758474
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM012
; CURRENT APPLICATION NUMBER: US/09/758,474
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1402
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 201
; LENGTH: 1409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (34)

; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (46)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1382)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1387)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-758-474-201

Query Match          43.1%; Score 22.4; DB 5; Length 1409;
Best Local Similarity 66.7%; Pred. NO. 12;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 2 cgcctgcctcactgcgggaccggtctaaagccggcccttaacccaacg 49
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1072 CCCTTCACACACTGATGGACACGCTACACAGGAGCGATACAAAGG 1025

RESULT 5
US-09-803-110-6363
; Sequence 6363, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 6363
; LENGTH: 837
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-803-110-6363

Query Match          42.7%; Score 22.2; DB 5; Length 837;
Best Local Similarity 69.8%; Pred. NO. 13;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 cgcctgcctcactgcgggaccggtctaaagccggcccttaacccaacg 44
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 96 cgcctgcctgattgcggcagggtgcgcgcctgctctcttgcgc 138

RESULT 6
US-09-803-110-4177
; Sequence 4177, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 4177
```

```
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-803-110-4177

Query Match          42.7%; Score 22.2; DB 5; Length 1395;
Best Local Similarity 69.8%; Pred. No. 14;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 cgttcgctcactgcgggaccggtctaaagccggccccccttaacc 44
||||| ||| | ||||| | | ||| | ||| |
Db 410 cgttcgctcgtcgccgaccgggtgacggcggaactgatcc 452

RESULT 7
US-09-764-902-493
; Sequence 493, Application US/09764902
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT213
; CURRENT APPLICATION NUMBER: US/09/764,902
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2318
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 493
; LENGTH: 3157
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (20)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (40)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2469)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (3138)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (3157)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-902-493

Query Match          42.7%; Score 22.2; DB 5; Length 3157;
Best Local Similarity 69.8%; Pred. No. 16;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3 gcttcgctcactgcgggaccggtctaaagccggccccccttaacca 45
||||| ||| | ||||| | | ||||| | ||| |
Db 1610 gcttcctcctcgtcgggccggtctaaacccggccgcagcagca 1652

RESULT 8
US-09-764-902-8896/c
; Sequence 8896, Application US/60278232
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using

; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; FILE REFERENCE: GX-0011 P
; CURRENT APPLICATION NUMBER: US/60/278,232
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 12,557
; SOFTWARE: PERL Program
; SEQ ID NO 8896
; LENGTH: 8416
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 402008.3
; NAME/KEY: unsure
; LOCATION: 367-594
; OTHER INFORMATION: a, t, c, g, or other
US-60-278-232-8896

Query Match          42.7%; Score 22.2; DB 6; Length 8416;
Best Local Similarity 69.8%; Pred. No. 18;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3 gcttcgctcactgcgggaccggtctaaagccggccccccttaacca 45
||||| ||| | ||||| | | ||||| | ||| |
Db 4325 GCTTCCTCTCCTGCTGGCGGCTAAACCCGGCGCGCAGCAGCA 4283

RESULT 9
US-09-764-902-2125/c
; Sequence 2125, Application US/09764902
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT213
; CURRENT APPLICATION NUMBER: US/09/764,902
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2318
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2125
; LENGTH: 12051
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-902-2125

Query Match          42.7%; Score 22.2; DB 5; Length 12051;
Best Local Similarity 69.8%; Pred. No. 19;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3 gcttcgctcactgcgggaccggtctaaagccggccccccttaacca 45
||||| ||| | ||||| | | ||||| | ||| |
Db 467 GCTTCCTCTCCTGCTGGCGGCTAAACCCGGCGCGCAGCAGCA 425

RESULT 10
US-09-764-902-2124/c
; Sequence 2124, Application US/09764902
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT213
; CURRENT APPLICATION NUMBER: US/09/764,902
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2318
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2124
; LENGTH: 18416
; TYPE: DNA
; ORGANISM: Homo sapiens
```

41.58: score 21.6: DB 5: Length 1145:

```

Best Local Similarity 75.0%; Pred. NO. 23;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 10 tcactgaggacggcgtacaaagcggcccttaacca 45
   || |||||  || ||||| ||||| |||||
Db 1048 tcgctgacggtcagccaaagctggccctgaacca 1083

RESULT 15
US-60-278-232-2701
; Sequence 2701, Application US/60278232
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; TITLE OF INVENTION: Polymorphisms Identified Thereby
; FILE REFERENCE: GX-0011 P
; CURRENT APPLICATION NUMBER: US/60/278,232
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 12,557
; SOFTWARE: PERL Program
; SEQ ID NO 2701
; LENGTH: 4202
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 198653.1
; NAME/KEY: unsure
; LOCATION: 2958-2987
; OTHER INFORMATION: a, t, c, g, or other
US-60-278-232-2701

```

```

Query Match 41.5%; Score 21.6; DB 6; Length 4202;
Best Local Similarity 75.0%; Pred. NO. 28;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 10 tcactgaggacggcgtacaaagcggcccttaacca 45
   || |||||  || ||||| ||||| |||||
Db 3372 tcgctgacggtcagccaaagctggccctgaacca 3407

```

Search completed: August 2, 2001, 16:55:08  
Job time: 117 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 18:38:56 ; Search time 1874.14 Seconds  
(without alignments)  
262.279 Million cell updates/sec

Title: US-09-701-626A-79  
Perfect score: 52  
Sequence: 1 tcgcttcgtcactgcggga.....ggcccttaacccaacgtta 52

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: gb\_est41:\*  
42: gb\_est42:\*  
43: gb\_est43:\*  
44: gb\_est44:\*  
45: gb\_est45:\*  
46: gb\_est46:\*  
47: gb\_est47:\*

44: en\_esthum10:\*  
45: en\_esthum11:\*  
46: en\_esthum12:\*  
47: en\_esthum13:\*  
48: en\_esthum14:\*  
49: en\_esthum15:\*  
50: en\_esthum16:\*  
51: en\_esthum17:\*  
52: en\_esthum18:\*  
53: en\_esthum19:\*  
54: en\_esthum20:\*  
55: en\_esthum21:\*  
56: en\_esthum22:\*  
57: en\_esthum23:\*  
58: en\_esthum24:\*  
59: en\_esthum25:\*  
60: en\_esthum26:\*  
61: en\_esthum27:\*  
62: en\_esthum28:\*  
63: en\_estin1:\*  
64: en\_estin2:\*  
65: en\_estin3:\*  
66: en\_estin4:\*  
67: en\_estin5:\*  
68: en\_estom1:\*  
69: en\_estom2:\*  
70: en\_estov1:\*  
71: en\_estov2:\*  
72: en\_estpl1:\*  
73: en\_estpl2:\*  
74: en\_estpl3:\*  
75: en\_estpl4:\*  
76: en\_estpl5:\*  
77: en\_estpl6:\*  
78: en\_estpl7:\*  
79: en\_estpl8:\*  
80: en\_estpl9:\*  
81: en\_estpl10:\*  
82: en\_estrol1:\*  
83: en\_estrol2:\*  
84: en\_estrol3:\*  
85: en\_estrol4:\*  
86: en\_estrol5:\*  
87: en\_estrol6:\*  
88: en\_estrol7:\*  
89: en\_estrol8:\*  
90: en\_estrol9:\*  
91: en\_estrol10:\*  
92: en\_estrol11:\*  
93: en\_estrol12:\*  
94: en\_estrol13:\*  
95: en\_estrol14:\*  
96: en\_estrol15:\*  
97: en\_estrol16:\*  
98: en\_estrol17:\*  
99: en\_estrol18:\*  
100: en\_estrol19:\*  
101: en\_estrol20:\*  
102: gb\_est25:\*  
103: gb\_est26:\*  
104: gb\_est27:\*  
105: gb\_est28:\*  
106: gb\_est29:\*  
107: gb\_est30:\*  
108: gb\_est31:\*  
109: gb\_est32:\*  
110: gb\_est33:\*  
111: gb\_est34:\*  
112: gb\_est35:\*  
113: gb\_est36:\*  
114: gb\_est37:\*  
115: gb\_est38:\*  
116: gb\_est39:\*

117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match					
C 1	23.8	45.8	553	21	A1520326	A1520326 LD40453.5	
C 2	23.8	45.8	563	21	A1517070	A1517070 GH27702.5	
C 3	23.8	45.8	991	219	CNS01320	AL102330 Drosophila	
C 4	23.6	45.4	1004	220	CNS027PB	AL184952 Tetraodon	
C 5	23.4	45.0	202	1	AA013941	AA013941 mh05C03.r	
C 6	23.4	45.0	593	121	CNS0337Z	BE853438 uw31f08.x	
C 7	23.4	45.0	879	221	CNS03C7E	AL242648 Tetraodon	
C 8	23.4	44.2	197	246	A2578481	AZ578481 23h08 Shoc	
C 9	22.8	43.8	342	163	BE074183	BE074183 RCI-BT056	
C 10	22.8	43.8	637	139	BE731519	BE731519 601569107	
C 11	22.8	43.8	815	174	BF316175	BG116175 602318362	
C 12	22.6	43.5	402	147	BF374103	BG116175 602318362	
C 13	22.6	43.5	509	168	BF726683	BF7374103 MR3-SN006	
C 14	22.6	43.5	515	228	AQ435257	BF726683 by10d08.Y	
C 15	22.6	43.5	924	153	BG390246	AQ435257 HS_5140_B	
C 16	22.4	43.1	347	116	AW438871	BG390246 602416203	
C 17	22.4	43.1	366	19	A1368921	AW438871 xtl3f06.x	
C 18	22.4	43.1	386	22	A1632338	A1368921 qv98c07.x	
C 19	22.4	43.1	395	20	A1434153	A1632338 t122904.x	
C 20	22.4	43.1	406	23	A1653217	A1434153 t136h07.x	
C 21	22.4	43.1	492	16	A1126769	A1653217 wp23h11.x	
C 22	22.4	43.1	518	116	AW440490	A1126769 qb85h08.x	
C 23	22.4	43.1	546	190	W19597	AW440490 hb91c05.x	
C 24	22.4	43.1	574	10	AA644914	W19597 zb36b12.r1	
C 25	22.4	43.1	584	144	BF107283	AA644914 v848404.r	
C 26	22.4	43.1	635	145	BF167553	BF107283 601824180	
C 27	22.4	43.1	643	249	AZ807425	BF167553 601774261	
C 28	22.4	43.1	716	10	AA671918	AZ807425 2M0070H18	
C 29	22.4	43.1	731	105	AL110412	AA671918 v111b02.r	
C 30	22.2	42.7	234	147	BF365198	AL110412 DKFPz434K	
C 31	22.2	42.7	373	104	AJ280017	BF365198 PM1-NT001	
C 32	22.2	42.7	433	225	AQ174469	AJ280017 4A3A-AAK-	
C 33	22.2	42.7	616	31	AV540656	AQ174469 HS_3215_A	
C 34	22.2	42.7	670	256	B17432	AV540656 AV540656	
C 35	22.2	42.3	519	222	FR0013145	B17432 345324.TVB	
C 36	22.2	42.3	541	20	A1463359	AL004395 F.rubrripe	
C 37	22.2	42.3	866	239	AZ207121	A1463359 mu75e01.x	
C 38	22.2	42.3	991	256	B11618	AZ207121 SP_0127_B	
C 39	21.8	41.9	310	108	AV019157	B11618 T8013-Sp6.3	
C 40	21.8	41.9	343	11	AA763260	AV019157 AV019157	
C 41	21.8	41.9	412	146	BF226128	AA763260 vV89b11.r	
C 42	21.8	41.9	412	115	AW412398	BF226128 uz47h06.Y	
C 43	21.8	41.9	435	165	BE290435	AW412398 uo78g08.Y	
C 44	21.8	41.9	458	1	AA266758	BE290435 60108886	
C 45	21.8	41.9	478	166	BE303443	A266758 mz93h06.r	
C 46	21.8	41.9	488	1	AA266758	BE303443 6010883837	

## ALIGNMENTS

RESULT	1	
LOCUS	AI520326	553 bp mRNA
DEFINITION	LD40453.5prime LD Drosophila melanogaster embryo pot2 Drosophila melanogaster cDNA clone LD40453 5prime, mRNA sequence.	EST 16-MAR-1999
ACCESSION	AI520326	
VERSION	AI520326.1	GI:4426180
KEYWORDS	EST.	
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	1 (bases 1 to 553)	
AUTHORS	Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein P., Lewis,S. and Rubin,G.M.	
TITLE	BDGP/HHMI Drosophila EST Project	

```

Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 404 row: E column: 5
High quality sequence stop: 462.
Location/Qualifiers
    1..553
     /organism="Drosophila melanogaster"
     /db_xref="taxon:7227"
     /clone_lib="LD40453"
     /clone_lib="JD Drosophila melanogaster embryo pot2"
     /sex="male and female"
     /dev_stage="0 to 24 hours mixed stage embryonic"
     /lab_host="XLI Blue"
     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2."
BASE COUNT      174 a   141 c   139 g   99 t
ORIGIN

Query Match          45.8%; Score 23.8; DB 21; Length 553;
Best Local Similarity 72.1%; Pred. NO. 72;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY  2  cgcttcgctcactgcggagccggcgtaaagcgccgcccttaacc 44
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   538 CGATTTTCGCACCAAGAGCGCAGGTAAAGCGCGCCATTAAACC 496

RESULT  2
A1517070/c
LOCUS       A1517070              563 bp             mRNA           EST
DEFINITION GH27702.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH27702 5prime, mRNA sequence.
ACCESSION   A1517070
VERSION     A1517070.1 GI:4420170
KEYWORDS    EST.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 563)
AUTHORS    Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
TITLE      BDGP/HHMI drosophila EST Project
JOURNAL     Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 277 row: A column: 2
            High quality sequence stop: 388.
            Location/Qualifiers
                1..563
                 /organism="Drosophila melanogaster"
                 /db_xref="taxon:7227"
                 /clone="GH27702"
                 /clone_lib="GH Drosophila melanogaster head pot2"
                 /sex="male and female"
                 /dev_stage="adult"
                 /lab_host="DH5 - alpha"
                 /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
XhoI; Sized fractionated cDNAs were directly ligated into
pOT2. Plasmid cDNA library."
BASE COUNT      176 a   145 c   140 g   102 t

```







```

Query Match      45.0%; Score 23.4; DB 221; Length 879;
Best Local Similarity 73.2%; Pred. No. 1.1e+02;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 5 ttgcctcactgcggaccggtctaaagccggcccttaacca 45
    ||||| | ||||| ||||| || | |||||
Db 232 TTGCACCTTTGGCGGACCGGCTAAACACCAACCACTTGAACCA 192

RESULT 8
A2578481/c
LOCUS      A2578481      197 bp      DNA      GSS      08-DEC-2000
DEFINITION 23h08 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
            sp. NGR234 genomic clone 23h08, DNA sequence.
ACCESSION  A2578481
VERSION    A2578481.1 GI:11605806
KEYWORDS
SOURCE    Rhizobium sp. NGR234.
ORGANISM  Rhizobium sp. NGR234
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE  1 (bases 1 to 197)
AUTHORS   Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
TITLE     Genetic snapshots of the Rhizobium species NGR234 genome
JOURNAL   GenomeBiology.com 1 (6), 0014.1-0014.7 (2000)
COMMENT   Contact: Virginie Viprey
            Laboratoire de Biologie Moleculaire des Plantes Superieures
            University of Geneva
            1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
            Tel: +44(0)1603450000
            Fax: +44(0)1603450045
            Email: virginie.viprey@bbrc.ac.uk
            Class: shotgun.
FEATURES   Location/Qualifiers
            1..197
                /organism="Rhizobium sp. NGR234"
                /strain="ANU265"
                /db_xref="taxon:394"
                /clone="23h08"
                /clone_lib="Shot-gun genomic library of Rhizobium strain
                ANU265."
                /note="Vector: M13; derivative strain of NGR234 cured of
                pNGR234a"
BASE COUNT  37 a 73 c 62 g 24 t 1 others
ORIGIN

```

```

Query Match      44.2%; Score 23; DB 246; Length 197;
Best Local Similarity 74.4%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 tcgcttcgctcactgcggaccggtctaaagccggccctc 39
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 166 TTGCTTCTCAAGCGGGGTTTCGGCGACGCGCGTCCGCT 128

```

```

RESULT 9
BE074183
LOCUS      BE074183      342 bp      mRNA      EST      09-JUN-2000
DEFINITION RCL-BT0567-301299-011-c04 BT0567 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BE074183
VERSION    BE074183.1 GI:8421866
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 342)
AUTHORS    Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

```

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=6t2-rc1-BT0567-301  
299-011-c04&t3=1999-12-30&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 342.  
Location/Qualifiers  
1..342  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BT0567"  
/dev\_stage="Adult"  
/note="Organ: breast; Vector: puc18; Site\_1: Smal; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
7,716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

FEATURES  
source

BASE COUNT 77 a 89 c 87 g 89 t  
ORIGIN

Query Match 43.8%; Score 22.8; DB 163; Length 342;  
Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 11 cactgcgggaccgctaaagccggcccttaaccacgctta 52  
 ||||| | ||||| ||||| ||||| ||||| ||||| |||||  
Db 230 CCCTGCGGTAGTGATATAAACCTGCACCAACCAACGTTA 271

RESULT 10  
BE735159  
LOCUS BE735159 637 bp mRNA EST 15-SEP-2000  
DEFINITION 601569107F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3844009 5',  
mRNA sequence.  
ACCESSION BE735159  
VERSION BE735159.1 GI:10149151  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 637)

REFERENCE  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cyapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLCM539 row: m column: 02  
 High quality sequence start: 3  
 High quality sequence stop: 589.  
 Location/Qualifiers

FEATURES

```

source
1. .637
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3844009"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA Synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 159 a 199 c 157 g 122 t

ORIGIN

```

Query Match 43.8%; Score 22.8; DB 139; Length 637;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 6 tcgctcactgcggacgcgtctaaagccgcgccttaacacaa 47
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 492 TCACTCACTACTGGAAGTCTCAAACTGGCCACTGAACAAA 533

RESULT 11
BG116175 815 bp mRNA EST 30-JAN-2001
LOCUS 602318362F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4418822 5',
DEFINITION mRNA sequence.
ACCESSION BG116175
VERSION BG116175.1 GI:12609681
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 815)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10153 row: k column: 15
High quality sequence stop: 733.
FEATURES
source
Location/Qualifiers
1. .815
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4418822"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
```

BASE COUNT 189 a 267 c 202 g 157 t

ORIGIN

```

Query Match 43.8%; Score 22.8; DB 174; Length 815;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 6 tcgctcactgcggacgcgtctaaagccgcgccttaacacaa 47
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 492 TCACTCACTACTGGAAGTCTCAAACTGGCCACTGAACAAA 533

RESULT 12
BF374103 402 bp mRNA EST 24-NOV-2000
LOCUS MR3-SN0066-020500-003-g01_1 SN0066 Homo sapiens cDNA, mRNA
DEFINITION sequence.
ACCESSION BF374103
VERSION BF374103.1 GI:11336232
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H.,
Brunstein, A., deOliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J., and
Simpson, A. J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3&t2=MR3-SN0066-
020500-003-g01.1&t3=2000-05-02&t4=1)
Seq primer: puc 18 forward
FEATURES
source
Location/Qualifiers
1. .402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="SN0066"
/dev_stage="Adult"
/note="Organ: stomach_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
```

BASE COUNT 92 a 103 c 86 g 121 t

ORIGIN

```

Query Match 43.5%; Score 22.6; DB 147; Length 402;
Best Local Similarity 68.9%; Pred. No. 2e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 8 gctcactgcggacgcgtctaaagccgcgccttaacacacgta 52
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2 GCTCATTCAGGACCGCTCATGTCCAGCGCCTAACCTTACCTTA 46

RESULT 13
BF726683/c 509 bp mRNA EST 05-JAN-2001
LOCUS BF726683
```

```

DEFINITION   by10d08.y1 Human Lens cDNA (Un-normalized, unamplified): BY Homo
ACCESSION    BF726683
VERSION      BF726683.1 GI:12042594
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE        Wistow,G.J., Bernstein,S., Behal,A. and Smith,D.
JOURNAL      NEIBANK: EST analysis and bioinformatics for ocular genomics
COMMENT      Invest. Ophthalmol. Vis. Sci. 41 (2000) In press
              Contact: Wistow G
              Section on Molecular Structure and Function
              National Eye Institute
              6/331, NIH, Bethesda, MD 20892-2740, USA
              Tel: 301 402 3452
              Fax: 301 496 0078
              Email: graeme@helix.nih.gov
              Plate: 10 row: d column: 08
              Seq primer: M13Rpl reverse primer (ABI).
              Location/Qualifiers
                1..509
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="by10d08"
                /clone_lib="Human Lens cDNA (Un-normalized, unamplified):
                BY"
                /tissue_type="Lens"
                /dev_stage="Adult"
                /lab_host="EMDH108"
                /note="Organ: Eye; Vector: pCMVSPORT6; Two human lenses
                from different adults (both approximately 40 years old)
                together yielded 20ug of total RNA and 150ng mRNA for cDNA
                library synthesis. A directionally cloned cDNA library in
                the pCMVSPORT6 vector was constructed at Life Technologies
                Plasmid System full details of the protocols of the SuperScript
                manufacturer's instruction manual
                (http://www.lifetech.com/). First strand synthesis was
                carried out using a Not I primer-adaptor
                [5'-pGACTGTTCTAGATCGGAGCGGCCCT(T)15-3']. Not I/blunt
                end inserts were cloned into the Not I/EcoR V sites in the
                vector. EST analysis was performed on the unamplified
                library at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT   95 a 144 c 156 g 113 t 1 others
ORIGIN
Query Match  43.5%; Score 22.6; DB 168; Length 509;
Best Local Similarity 68.9%; Pred. No. 2e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 4 cttcgtcactcgggacgcgttaaacgcccgccttaaccaaac 48
    ||||| 1 ||||| ||| 1 ||||| 1 |||
Db 95 CTCTCAGACTCGGGGACCTGCAGCACCCTCGCCCGCCAGTCTCTGAAC 51
    ||||| 1 ||||| ||| 1 ||||| 1 |||

RESULT 14
LOCUS       A0435257
DEFINITION HS_5140_B2_H04.SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=716 Col=8 Row=P, DNA sequence.
ACCESSION  A0435257
VERSION     A0435257.1 GI:4546596
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

```

```

TITLE        Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
JOURNAL      Hood,L.
MEDLINE      Sequence-tagged connectors: A sequence approach to mapping and
COMMENT      scanning the human genome
              Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
              95380589
              Contact: Mahairas GG, Wallace JC, Hood L
              High Throughput Sequencing Center
              University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
              Email: jwallace@u.washington.edu
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieter@dejong.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
              or from Resear h Genetics (info@resgen.com). BAC end Web Server:
              http://www.htsc.washington.edu
              Plate: 716 row: P column: 8
              Seq primer: SP6
              Class: BAC ends
              High quality sequence stop: 515.
              Location/Qualifiers
                1..515
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate=716 Col=8 Row=P"
                /clone_lib="RPCI-11 Human Male BAC Library"
                /sex="male"
                /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                Male blood DNA was isolated from one randomly chosen donor
                and partially digested with a combination of EcoRI and
                EcoRI Methylase. Size selected DNA was cloned into the
                pBACe3.6 vector at EcoRI sites"
BASE COUNT   154 a 88 c 91 g 179 t 3 others
ORIGIN
Query Match  43.5%; Score 22.6; DB 228; Length 515;
Best Local Similarity 68.9%; Pred. No. 2e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 8 gctcactcgggacgcgttaaacgcccgccttaaccaaacgtta 52
    ||||| 1 ||||| ||| 1 ||||| 1 |||
Db 263 GATTTCTGCTGACAGCTTTAAAGCAGCCCTCTTAAACAAAAGATA 307
    ||||| 1 ||||| ||| 1 ||||| 1 |||

RESULT 15
LOCUS       BG390246
DEFINITION BG390246 924 bp mRNA EST 12-MAR-2001
            602416203F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4524648 5',
            mRNA sequence.
ACCESSION  BG390246
VERSION     BG390246.1 GI:13283694
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              NIH-MGC http://mgc.nci.nih.gov/.
              National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
              Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-remail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov

```

```

Plate: LLAM10429 row: e column: 01
High quality sequence stop: 718.
Location/Qualifiers
1. .924
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4524848"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT      265 a      186 c      162 g      311 t
ORIGIN

```

```

Query Match      43.5%; Score 22.6; DB 153; Length 924;
Best Local Similarity 75.7%; Pred. No. 2.1e+02;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 16 cgggaccgctaaagcggcccttaaccaaagttta 52
    ||||| ||| || | ||| | | ||||| |
Db 814 CGGACCCGCTTAACCGGTCCTTAACCCAAACGTAA 850

```

Search completed: August 2, 2001, 18:39:11  
Job time: 6361 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 18:06:17 ; Search time 1350.43 Seconds  
(without alignments)  
595.605 Million cell updates/sec

Title: US-09-701-626A-79  
Perfect score: 52  
Sequence: 1 tcgtcttcgtcactgcggga.....ggcccttaaccaaacgtta 52

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba1.\*
- 2: gb\_ba2.\*
- 3: gb\_ba3.\*
- 4: gb\_in1.\*
- 5: gb\_in2.\*
- 6: gb\_in3.\*
- 7: gb\_om.\*
- 8: gb\_ov.\*
- 9: gb\_pat1.\*
- 10: gb\_pat2.\*
- 11: gb\_ph.\*
- 12: gb\_p11.\*
- 13: gb\_p12.\*
- 14: gb\_p13.\*
- 15: gb\_p14.\*
- 16: em\_ba1.\*
- 17: em\_ba2.\*
- 18: em\_fun.\*
- 19: em\_htgo\_hum.\*
- 20: em\_htgo\_inv.\*
- 21: em\_htgo\_rod.\*
- 22: em\_htg\_hum1.\*
- 23: em\_htg\_hum2.\*
- 24: em\_htg\_hum3.\*
- 25: em\_htg\_hum4.\*
- 26: em\_htg\_hum5.\*
- 27: em\_htg\_hum6.\*
- 28: em\_htg\_hum7.\*
- 29: em\_htg\_hum8.\*
- 30: em\_htg\_inv1.\*
- 31: em\_htg\_inv2.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_rod.\*
- 34: em\_hum1.\*
- 35: em\_hum2.\*
- 36: em\_hum3.\*
- 37: em\_hum4.\*
- 38: em\_hum5.\*
- 39: em\_hum6.\*
- 40: em\_hum7.\*
- 41: em\_in.\*
- 42: em\_om.\*
- 43: em\_or.\*

- 44: em\_ov.\*
- 45: em\_pat.\*
- 46: em\_ph.\*
- 47: em\_pl.\*
- 48: em\_ro.\*
- 49: em\_sts.\*
- 50: em\_sy.\*
- 51: em\_un.\*
- 52: em\_vi.\*
- 53: gb\_sts1.\*
- 54: gb\_sts2.\*
- 55: gb\_sts3.\*
- 56: gb\_sy.\*
- 57: gb\_un.\*
- 58: gb\_vil.\*
- 59: gb\_v12.\*
- 60: gb\_htg1.\*
- 61: gb\_htg2.\*
- 62: gb\_htg3.\*
- 63: gb\_htg4.\*
- 64: gb\_htg5.\*
- 65: gb\_htg6.\*
- 66: gb\_htg7.\*
- 67: gb\_htg8.\*
- 68: gb\_htg9.\*
- 69: gb\_htg10.\*
- 70: gb\_htg11.\*
- 71: gb\_htg12.\*
- 72: gb\_htg13.\*
- 73: gb\_htg14.\*
- 74: gb\_htg15.\*
- 75: gb\_htg16.\*
- 76: gb\_htg17.\*
- 77: gb\_htg18.\*
- 78: gb\_htg19.\*
- 79: gb\_htg20.\*
- 80: gb\_htg21.\*
- 81: gb\_htg22.\*
- 82: gb\_htg23.\*
- 83: gb\_htg24.\*
- 84: gb\_htg25.\*
- 85: gb\_pr1.\*
- 86: gb\_pr2.\*
- 87: gb\_pr3.\*
- 88: gb\_pr4.\*
- 89: gb\_pr5.\*
- 90: gb\_pr6.\*
- 91: gb\_pr7.\*
- 92: gb\_pr8.\*
- 93: gb\_pr9.\*
- 94: gb\_rol.\*
- 95: gb\_rol2.\*
- 96: gb\_in4.\*
- 97: gb\_pr10.\*
- 98: em\_ba3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
C 1	24.2	46.5	1712	3 PSELASA	M20982 P.aeruginos
C 2	24.2	46.5	3742	3 PAU68175	U68175 Pseudomonas
C 3	23.8	45.8	4943	63 AC014389	AC014389 Drosophil
C 4	23.8	45.8	294218	4 AE003430	AE003430 Drosophil
C 5	23.6	45.4	9581	3 PAE223604	AJ223604 Pseudomon
C 6	23.2	44.6	68596	69 AC025118	AC025118 Homo sapi
C 7	23.2	44.6	181052	91 CNS01DWP	AL137818 Human chr
C 8	23	44.2	33615	2 BSU11039	U11039 Bacillus su

```
9 23 44.2 120766 85 AC004150
10 23 44.2 208780 2 BSUB0009
11 23 44.2 218864 68 AC023855
12 23 44.2 233780 2 BSUB0010
13 22.8 43.8 172569 79 AL353608
14 22.6 43.5 1142 12 AF033487
15 22.6 43.5 45257 76 AC084177
16 22.6 43.5 95255 88 AF051976
17 22.6 43.5 133075 78 AF205589
18 22.6 43.5 163608 69 AC026081
19 22.6 43.5 173749 66 AC020567
20 22.4 43.1 1233 94 MNCERBA1
21 22.4 43.1 1240 94 MNCERBA3
22 22.4 43.1 1357 94 MNCERBA2
23 22.4 43.1 1440 94 MNCERBA
24 22.4 43.1 167294 63 AC013814
25 22.4 43.1 179023 73 AC068768
26 22.2 42.7 173 9 AR095503
27 22.2 42.7 303 58 AB026094
28 22.2 42.7 303 58 AB026096
29 22.2 42.7 303 58 AB026097
30 22.2 42.7 303 58 AB026098
31 22.2 42.7 775 58 AF075212
32 22.2 42.7 822 58 AF058736
33 22.2 42.7 117620 82 AL590618
34 22.2 42.7 128396 71 AC034262
35 22.2 42.7 132805 92 HS339A18
36 22.2 42.7 137604 67 AC022275
37 22.2 42.7 148204 74 AC069400
38 22.2 42.7 164630 66 AC020632
39 22.2 42.7 171744 83 CNS01DM9
40 22.2 42.7 180238 74 AC073688
41 22.2 42.7 182741 65 AC018910
42 22.2 42.7 184825 90 AL391380
43 22.2 42.7 195705 68 AC024270
44 22.2 42.7 224592 75 AC074209
45 22 42.3 591 54 ECU90592
```

## ALIGNMENTS

```
RESULT 1
PSLASA/c 1712 bp DNA BCT 26-APR-1993
LOCUS P.aeruginosa structural elastase protein (lasA) gene, complete cds.
DEFINITION M20982.1 GI:151323
KEYWORDS lasA gene; structural elastase protein.
SOURCE P.aeruginosa DNA, clone pPS1816.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 1712)
Schad,P.A. and Iglewski,B.H.
Nucleotide sequence and expression in Escherichia coli of the
Pseudomonas aeruginosa lasA gene
J. Bacteriol. 170, 2784-2789 (1988)
88227864
COMMENT This protein will react with elastase antibodies but shows no
elastase activity.
FEATURES
Location/Qualifiers
1..1712
/organism="Pseudomonas aeruginosa"
/db_xref="taxon:287"
148..152
RBS /note="S-D ribosomal binding site (put.): putative"
159..1292
CDS /note="structural elastase"
/codon_start=1
/transl_table=11
/protein_id="AAA25873.1"
/db_xref="GI:151324"
```

```
/translation="MOHKRRAMASPRSPFLFVLALAVGTTANAHDDGLPAFRYSAE
LLGQLQPSVALPLNDDLLFYGRDAEAFDEALAYLALNAPALRDKSEYLHWGYSYIN
PKVLLTLVMQSGPLGADERLAALGLRLSAKRGFDQVRDLVQLSRRYKGFEEYQ
LROAAARKAVGEDGLNAAASALLGLREGAKVSARVARQSAAYACTFORLFGTFAAE
LLOPSNRVAROLOAKAALAPPSNLMLPWRQGYSHQPNGAHFEGSGYPYSSFDASYD
WPRWGSATYSVVAAHAGTVRVLRSQCVRVTHPSGNAATNYHMDQIQVSNQGVASDTK
LGVIAGNITALCEGGSSTGPHLHFSLLYNGAFVSLQ...SFGP"
BASE COUNT 276 a 596 c 561 g 279 t
ORIGIN 17 bp upstream of SmaI site.

Query Match 46.5%; Score 24.2; DB 3; Length 1712;
Best Local Similarity 78.4%; Pred. No. 1.7e+02;
Matches 29; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 cgccttcgctcactgcggcgaccgtaagccgccc 38
|| ||||| |||| ||||| ||||| ||||| |||
Db 1463 CGTTTCGCTCGCGGCGGCGGCTGGAGCCGCGCC 1427

RESULT 2
PAU68175/c
LOCUS PAU68175 3742 bp DNA BCT 06-APR-1998
DEFINITION Pseudomonas aeruginosa staphylytic protease preproenzyme LasA
(lasA) gene, complete cds.
ACCESSION U68175
VERSION 1
KEYWORDS Pseudomonas aeruginosa.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 3742)
Gustin,J.K., Kessler,E. and Ohman,D.E.
A substitution at His-120 in the LasA protease of Pseudomonas
aeruginosa blocks enzymatic activity without affecting propeptide
processing or extracellular secretion
J. Bacteriol. 178 (22), 6608-6617 (1996)
97086629
REFERENCE 2 (bases 1 to 3742)
Gustin,J.K., Kessler,E. and Ohman,D.E.
Direct Submission
Submitted (27-AUG-1996) Microbiology and Immunology, University of
Tennessee and Veterans Affairs Medical Center, 858 Madison Ave.,
Room 101, Memphis, TN 38163, USA
FEATURES
Source
1..3742
/organism="Pseudomonas aeruginosa"
/strain="FRD1"
/db_xref="taxon:287"
/map="55 min."
658..660
misc_feature /note="transcriptional start site"
evidence=not_experimental
763..2030
gene /gene="lasA"
763..767
RBS /gene="lasA"
774..866
sig_peptide /gene="lasA"
774..2030
CDS /gene="lasA"
/codon_start=1
/transl_table=11
/product="staphylytic protease preproenzyme LasA"
/protein_id="AAC12656.1"
/db_xref="GI:1549397"
/translation="MOHKRRAMASPRSPFLFVLALAVGTTANAHDDGLPAFRYSAE
LLGQLQPSVALPLNDDLLFYGRDAEAFDEALAYLALNAPALRDKSEYLHWGYSYIN
PKVLLTLVMQSGPLGADERLAALGLRLSAKRGFDQVRDLVQLSRRYKGFEEYQ
LROAAARKAVGEDGLNAAASALLGLREGAKVSARVARQSAAYACTFORLFGTFAAE
LLOPSNRVAROLOAKAALAPPSNLMLPWRQGYSHQPNGAHFEGSGYPYSSFDASYD
WPRWGSATYSVVAAHAGTVRVLRSQCVRVTHPSGNAATNYHMDQIQVSNQGVASDTK"
```

mat_peptide	LVYAGNINTALCEGSSGTPHLHFSILLYNGAPVSLQGSFGPYRINVGTSNTYNDGCR RYFYNQSGAGTTHCAFRPLYNPGAL" /gene="lasa" /note="staphylolytic protease" /product="Lasa" /b39..1841 /gene="lasa" /b39..1841									
misc_feature										
CDS	activity" complement(2144..2935) /note="ORF1" /codon_start=1 /transl_table=11 /protein_id="AAC12657.1" /db_xref="GI:1549398" /translation="MSLSPADLKQLQAANKLNPNGLTAQLTSLTGTPIEYGLKLPRL KVEKLGSVIEGALLKVARSAATLQENASGOPANRLHTLGAAVSGVGFGFGLPLML VEVPVTGIIIFRSIADISEGESIRDMDTLMACIEVFALGGKASDDASESGYAVR TALAOOVKAAPDLGKDGSKKAAPMLALVRKVAERLGVYSEKLAQLVIPVIGALG GAINTVFMHFQAMARGHFVVRRLRHHGKDVVRDAYALLPRG"									
BASE COUNT	631 a	1236 c	1278 g	597 t						
ORIGIN										
Query Match	46.5%	Score 24.2:	DB 3:	Length 3742;						
Best Local Similarity	78.4%	Pred. No. 1.4e+02:								
Matches	29;	Conservative	0;	Mismatches	8;	Indels	0;	Gaps	0;	
QY	2	cgcttcgctcactgcgggacgcgctaaagcgcggcccc 38								
Db	2076	CGTTTCGTCGCGCGGCGCGGCTGGAGCGCGGCCCC 2040								
RESULT	3									
AC014389/c										
LOCUS	AC014389 4943 bp DNA HTG 16-NOV-1999									
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.									
ACCESSION	AC014389									
VERSION	AC014389.1 GI:6436946									
KEYWORDS	HTG; HTGS_PHASE2.									
SOURCE	fruit fly.									
ORGANISM	Drosophila melanogaster									
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.									
AUTHORS	1 (bases 1 to 4943)									
TITLE	Adams,M. and Venter,J.C.									
JOURNAL	Direct Submission									
COMMENT	Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA This sequence was identified as CDN:10214250 by the submitter. For further information on this sequence e-mail to flycelera.com. * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.									
FEATURES	Location/Qualifiers									
source	1..4943 /organism="Drosophila melanogaster" /db_xref="taxon:7227"									
BASE COUNT	1518 a	1061 c	1096 g	1268 t						
ORIGIN										
Query Match	45.8%	Score 23.8:	DB 63:	Length 4943;						
Best Local Similarity	72.1%	Pred. No. 1.8e+02:								
Matches	31;	Conservative	0;	Mismatches	12;	Indels	0;	Gaps	0;	
QY	2	cgcttcgctcactgcgggacgcgctaaagcgcggcccttaacc 44								
Db	1092	CGATTCGCGACCAAGAGCGAGGCTAAGGCGCGCCATTAAAC 1050								





```

VERSION      AJ223604.1  GI:4210822
KEYWORDS
SOURCE       Pseudomonas aeruginosa.
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
              Pseudomonas.
REFERENCE    1 (bases 1 to 9581)
AUTHORS      Laraki,N., Galleni,M., Thamm,I., Riccio,M.L., Amicosante,G.,
              Frere,J.M. and Rossolini,G.M.
TITLE        Structure of In31, a blaIMP-containing Pseudomonas aeruginosa
              integron phyletically related to In5, which carries an unusual
              array of gene cassettes
JOURNAL      Antimicrob. Agents Chemother. 43 (4), 890-901 (1999)
MEDLINE      99216901
REFERENCE    2 (bases 1 to 9581)
AUTHORS      Galleni,M.G.
TITLE        Direct Submission
JOURNAL      Submitted (20-JAN-1998) Galleni M.G., Centre for Protein
              Engineering (CIP), University of Liege, B6 Sart Tilman, Liege,
              Liege B4000, Belgium
FEATURES     Location/Qualifiers
              source          1..9581
                           /organism="Pseudomonas aeruginosa"
                           /db_xref="taxon:287"
              repeat_region   117..140
                           /note="IRi"
                           /rpt_type="INVERTED"
              misc_feature    117..1467
                           /note="5' conserved element"
              gene            complement(319..1332)
                           /gene="int11"
              CDS              complement(319..1332)
                           /gene="int11"
                           /codon_start=1
                           /transl_table=11
                           /protein_id="CA11470.1"
                           /db_xref="GI:4210823"
                           /db_xref="SPTREMBL:Q9R819"
                           /translation="MKTATAPLPLRSVKVLDQLRIRIRYLHYSLRTEQAYVHVWRAF
IRFHGRPATLGSSEVEAFSLWNERKVSVSTHQALAAALLFFYKVLCTDLPLWQ
EIGRPPRRRLPVLTPDVRILGLFEGEHLFAQLLYGTGMRISEGLQLRVKDLDF
DHGTTIVREGKSKDRALMLPESLAPTLREQLSRANWMLKDQASGRSGVALPDALER
KYPRAGHSWPWPVWFAQHTHSTDPRSGVYVRHHMVDQTFQRAFKRAVEAGGITKATP
HTLRHSFATALLRSGLDTRTQVODLLGSDVSTMTIYTHVLKVGAGVRSPLDALPLPT
SER"
              mRNA            1468..2346
                           /gene="bla-imp"
              gene            1468..2346
                           /gene="bla-imp"
              CDS              1485..2225
                           /gene="bla-imp"
                           /codon_start=1
                           /transl_table=11
                           /protein_id="CA11471.1"
                           /db_xref="GI:4210824"
                           /translation="MKSLSVFIFLFCSTIATAESLPDLKIEKLDGCVVYVHTSFEEVN
GMGWPGKGLVVLNAEAYLIDTPTAKDTKLVTFVFERGYKIKGSISHSFSDSTG
GIENLNSIETFYAGSELTNELKKDGRQVATNSFGVNYWLVKNKIEFYVPGPGHTPD
NVWVLPERRKILFPGCFIKPYGLGNLGDANI EAWPKSAKLLSKYKAKLVVYPSHSEV
GDASLLKLTLEQAVKGLNESKPKSPN"
              misc_feature    2262..2346
                           /gene="bla-imp"
                           /note="59 bp element"
              mRNA            2347..2984
                           /gene="aacA4"
              gene            2347..2984
                           /gene="aacA4"
              CDS              2371..2925
                           /gene="aacA4"
                           /codon_start=1
                           /transl_table=11
                           /protein_id="CA11472.1"
                           /db_xref="GI:4210825"

/db_xref="SPTREMBL:Q51410"
/translation="MTNSTDSVTLRLMTEHDLAMLYEWNLRSHIVEMWGGEEAPTLLA
DVQEQYLPYSYLAQESVTPYIAMLNGEPYIAQSYVALGSDGDMWEEETDPGVRGIDQL
LANASOLKGLGTKLVRALVELLFPNDPEVTKIQTDPSNLRATRCYEKAGFERGTV
TTPDGPAYVWVQTRQAFERTSDA"
2920..2984
/gene="aacA4"
/note="59 bp element"
2985..3715
/gene="catB6"
2985..3715
/gene="catB6"
3018..3650
/gene="catB6"
/codon_start=1
/transl_table=11
/protein_id="CA11473.1"
/db_xref="GI:4210826"
/db_xref="SPTREMBL:Q9R818"
/translation="MENYFDSFPFKGKLLSEQVTNRNIKVGYSYVSGYYHIGHSFDDCA
RYLLPDRDQVDKLIIGSFCSIGSGASFIMAGNQHGRHDWVTSEPFYMOEPEAFSSST
DAFOKAGDITVGNQDVMIGSEAMIPGKIGDGAIVGSRSLVTRDVEPYTIIIGNPAKO
IKKRFSDDEISLLMEMEWNWPLDKIKTAMPLLCSSDIFGLHRRHWGRTAV"
3645..3715
/gene="catB6"
/note="59 bp element"
3716..4404
/gene="orf1"
3716..4404
/gene="orf1"
3727..4341
/gene="orf1"
/codon_start=1
/transl_table=11
/protein_id="CA11474.1"
/db_xref="GI:4210827"
/db_xref="SPTREMBL:Q9ZEHL"
/translation="MDVDLQSIGIQTVQEALEKGLIFNEKTKTDDFNKACIHLIVLL
VDAQCFDGRSGWCTSVFLSHTALEEVAKEVGLYRRGCKIGKARGKDKLPFNHGEKHR
MALTLPVFKSRLEELALGKKECAELKKEAAGFRNRSLSYFSNENGQEVTPANVV
SQNRARELLLEAADDRLVGYTNHTGTILEAKINEIFSVAHS"
4335..4404
/gene="orf1"
/note="59 bp element"
4405..4940
/gene="gacE2"
4405..4940
/gene="gacE2"
4507..4839
/gene="gacE2"
4507..4839
/gene="gacE2"
/codon_start=1
/transl_table=11
/protein_id="CA11475.1"
/db_xref="GI:4210828"
/db_xref="SPTREMBL:Q9ZEHO"
/translation="MKNWLFATAIIFEVIATATSAKSSGEGFTRLVPSFIIVAGYAAAF
YFLSLTKSPIPGIAYAVWSGLGIVLTAVWVLRGQKLDWMGFGVGVGTIISGVAVLN
LLSKASVH"
4849..4936
/gene="gacE2"
/note="59 bp element"
4941..6961
/note="3' conserved element"
5045..5392
/gene="gacEdelta 1"
5045..5392
/gene="gacEdelta 1"
/codon_start=1
/transl_table=11
/protein_id="CA11476.1"
/db_xref="GI:4210829"

```



\* 8939 9038: gap of 100 bp  
\* 9039 9782: contig of 744 bp in length  
\* 9783 9882: gap of 100 bp  
\* 9883 10608: contig of 726 bp in length  
\* 10609 10708: gap of 100 bp  
\* 10709 11442: contig of 734 bp in length  
\* 11443 11542: gap of 100 bp  
\* 11543 12259: contig of 717 bp in length  
\* 12260 12359: gap of 100 bp  
\* 12360 13080: contig of 721 bp in length  
\* 13081 13180: gap of 100 bp  
\* 13181 13904: contig of 724 bp in length  
\* 13905 14004: gap of 100 bp  
\* 14005 14734: contig of 730 bp in length  
\* 14735 14834: gap of 100 bp  
\* 14835 15563: contig of 729 bp in length  
\* 15564 15663: gap of 100 bp  
\* 15664 16375: contig of 712 bp in length  
\* 16376 16475: gap of 100 bp  
\* 16476 17198: contig of 724 bp in length  
\* 17200 17299: gap of 100 bp  
\* 17300 18031: contig of 732 bp in length  
\* 18032 18131: gap of 100 bp  
\* 18132 18874: contig of 743 bp in length  
\* 18875 18974: gap of 100 bp  
\* 18975 19693: contig of 725 bp in length  
\* 19700 19799: gap of 100 bp  
\* 19800 20524: contig of 725 bp in length  
\* 20525 20624: gap of 100 bp  
\* 20625 21333: contig of 709 bp in length  
\* 21334 21433: gap of 100 bp  
\* 21434 22172: contig of 739 bp in length  
\* 22173 22272: gap of 100 bp  
\* 22273 23005: contig of 733 bp in length  
\* 23006 23105: gap of 100 bp  
\* 23106 23848: contig of 743 bp in length  
\* 23849 23948: gap of 100 bp  
\* 23949 24676: contig of 728 bp in length  
\* 24677 24776: gap of 100 bp  
\* 24777 25494: contig of 718 bp in length  
\* 25495 25594: gap of 100 bp  
\* 25595 26332: contig of 738 bp in length  
\* 26333 26432: gap of 100 bp  
\* 26433 27173: contig of 741 bp in length  
\* 27174 27273: gap of 100 bp  
\* 27274 28036: contig of 763 bp in length  
\* 28037 28136: gap of 100 bp  
\* 28137 28855: contig of 719 bp in length  
\* 28856 28955: gap of 100 bp  
\* 28956 29688: contig of 733 bp in length  
\* 29689 29788: gap of 100 bp  
\* 29789 30535: contig of 747 bp in length  
\* 30536 30635: gap of 100 bp  
\* 30636 31368: contig of 733 bp in length  
\* 31369 31468: gap of 100 bp  
\* 31469 32201: contig of 733 bp in length  
\* 32202 32301: gap of 100 bp  
\* 32302 33012: contig of 711 bp in length  
\* 33013 33112: gap of 100 bp  
\* 33113 33853: contig of 741 bp in length  
\* 33854 33953: gap of 100 bp  
\* 33954 34674: contig of 721 bp in length  
\* 34675 34774: gap of 100 bp  
\* 34775 35496: contig of 722 bp in length  
\* 35497 35596: gap of 100 bp  
\* 35597 36334: contig of 738 bp in length  
\* 36335 36434: gap of 100 bp  
\* 36435 37172: contig of 738 bp in length  
\* 37173 37272: gap of 100 bp  
\* 37273 37998: contig of 726 bp in length  
\* 37999 38098: gap of 100 bp  
\* 38099 38820: contig of 722 bp in length  
\* 38821 38920: gap of 100 bp

\* 38921 39621: contig of 701 bp in length  
\* 39622 39721: gap of 100 bp  
\* 39722 40427: contig of 706 bp in length  
\* 40428 40527: gap of 100 bp  
\* 40528 41260: contig of 733 bp in length  
\* 41261 41360: gap of 100 bp  
\* 41361 42100: contig of 740 bp in length  
\* 42101 42200: gap of 100 bp  
\* 42201 42932: contig of 732 bp in length  
\* 42933 43032: gap of 100 bp  
\* 43033 43747: contig of 715 bp in length  
\* 43748 43847: gap of 100 bp  
\* 43848 44577: contig of 730 bp in length  
\* 44578 44677: gap of 100 bp  
\* 44678 45392: contig of 715 bp in length  
\* 45393 45492: gap of 100 bp  
\* 45493 46203: contig of 717 bp in length  
\* 46210 46309: gap of 100 bp  
\* 46310 47028: contig of 719 bp in length  
\* 47029 47128: gap of 100 bp  
\* 47129 47852: contig of 724 bp in length  
\* 47853 47952: gap of 100 bp  
\* 47953 48674: contig of 722 bp in length  
\* 48675 48774: gap of 100 bp  
\* 48775 49504: contig of 730 bp in length  
\* 49505 49604: gap of 100 bp  
\* 49605 50333: contig of 729 bp in length  
\* 50334 50433: gap of 100 bp  
\* 50434 51170: contig of 737 bp in length  
\* 51171 51270: gap of 100 bp  
\* 51271 52001: contig of 731 bp in length  
\* 52002 52101: gap of 100 bp  
\* 52102 52838: contig of 737 bp in length  
\* 52839 52938: gap of 100 bp  
\* 52939 53671: contig of 733 bp in length  
\* 53672 53771: gap of 100 bp  
\* 53772 54533: contig of 762 bp in length  
\* 54534 54633: gap of 100 bp  
\* 54634 55354: contig of 721 bp in length  
\* 55355 55454: gap of 100 bp  
\* 55455 56183: contig of 729 bp in length  
\* 56184 56283: gap of 100 bp  
\* 56284 57008: contig of 725 bp in length  
\* 57009 57108: gap of 100 bp

Query Match 44.6%; Score 23.2; DB 69; Length 68596;  
Best Local Similarity 70.5%; Pred. No. 1.7e+02;  
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 9 ctcaactcgaggacgcgctaaagccgcgccttaaccacagttta 52  
||||||| ||| ||||| ||| ||||| ||| |||||  
Db 54835 CTCACGTGTTCCGGCTACAGTTGCCCTTGATGAGTGTTA 54878

RESULT 7

CNS01DWP

LOCUS

DEFINITION

CNS01DWP 181052 bp DNA PRI 25-MAY-2000  
Human chromosome 14 DNA sequence \*\*\* IN PROGRESS \*\*\* BAC R-102C24  
of library RPCI-11 from chromosome 14 of Homo sapiens (Human),  
complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 181052)  
Genoscope.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (23-MAY-2000) to the EMBL/GenBank/DBJ databases  
On Feb 20, 2000 this sequence version replaced gi:6982193.  
IMP(TANT: This sequence is unfinished and does not necessarily

represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continue. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. . . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the SP6 end.

Upstream BAC (overlapping the T7 end) : R-317N8  
Downstream BAC (overlapping the SP6 end) : R-355C3

-----  
Overall quality chart :

Range	: bases
0	:
1 - 9	: 93
10 - 19	: 320
20 - 29	: 1291
30 - 39	: 3128
40 - 49	: 5327
50 - 59	: 12776
60 - 69	: 30990
70 - 79	: 58042
80 - 89	: 42197
90 - 99	: 26888

Percentage of bases with a quality value >= 40 : 97 %

FEATURES  
Source

1..181052 Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone\_lib="RPC1-11"  
/clone="R-102C24"

BASE COUNT 52464 a 32038 c 33821 g 62729 t

ORIGIN

Query Match 44.6%; Score 23.2; DB 91; Length 181052;  
Best Local Similarity 70.5%; Pred. No. 1.4e+02;  
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
  
QY 9 ctcaactgcggagccgctaaagccggcccttaaccaaagctta 52  
||||| | | | | | | | | | | | | | | | | | | | |  
Db 116656 CTCACGTGTGTCGCGCTACAGTTTGCCCTTGATGAGTGTTA 116699

RESULT 8  
BSU11039  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
AUTHORS  
TITLE  
Direct Submission

BSU11039 33615 bp DNA BCT 06-JAN-1996  
Bacillus subtilis w168 polyketide synthase (pkx and pksofx6)  
genes, complete cds.  
U11039 M97902  
U11039.1 GI:602656  
Bacillus subtilis.  
Bacillus subtilis  
Bacteria; Firmicutes; Bacillus/Clostridium group;  
Bacillus/Staphylococcus group; Bacillus.  
1 (bases 20000 to 33615)  
Scotti,C., Piatelli,M., Cuzzoni,A., Perani,P., Tognoni,A., Grandi,G.,  
Galizzi,A. and Albertini,A.M.  
A Bacillus subtilis large ORF coding for a polypeptide highly  
similar to polyketide synthases  
Gene 130 (1), 65-71 (1993)  
93345824  
2 (bases 1 to 20000)  
Albertini,A.M., Caramori,T., Scoffone,F., Scotti,C. and Galizzi,A.  
Sequence around the 159 degree region of the Bacillus subtilis  
genome: the pkx locus spans 33.6 kb  
Microbiology 141 (Pt 2), 299-309 (1995)  
95219083  
3 (bases 1 to 33615)  
Albertini,A.M.  
Direct Submission

JOURNAL Submitted (20-JUN-1994) Alessandra M. Albertini, Dipartimento di  
Genetica e Microbiologia, Universita di Pavia, 207, via  
Abbategrasso, Pavia I-27100, Italy  
COMMENT On Dec 17, 1994 this sequence version replaced gi:528989.  
FEATURES  
Source  
1..33615  
/organism="Bacillus subtilis"  
/strain="168"  
/db\_xref="taxon:1423"  
323..369  
RBS  
785..789  
CDS  
795..2045  
/note="pksofx1; similar to Rhizobium leguminosarum  
nodulation protein E, Swiss-Prot Accession Number P04684,  
and to E. coli 3-oxoacyl-(acyl-carrier protein) synthase  
I, Swiss-Prot Accession Number P14926"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AA085139.1"  
/db\_xref="GI:602657"  
/translation="MTKCNLPVVVTVGVGTASIGOGKEDFASLLSGRHAFDVMKRS  
GRKDSRFIGAELASLSYDPRLSKMLRKASPSRRAALVLTLEWEEAEADDSERI  
GLVGSNPFQRENFYQDRSGFISPAYGLSDSLGICITDQFGITGLATV  
GASASGQQLAVIHAIQQVLSGEVDTICIALGMLDLSYMECEALRGAMGDTKYADE  
PENACRPFDQNRDGFYIGSCGALVIERKETALRGKLPYAALSGWSIKLDGNNRDP  
SLEGEITHVQKALERARLLPEDIDYINPHGTGTFIGDETELKALRACRLSHAYINATK  
SITGHLSAAGIVEIISVLLQMKKSAHPSRNLDPIDDSFHHVNEKSIYRIKNALS  
LSMGFGMNTAVCIQNIKCGES"  
1961..1989  
terminator  
-35\_signal  
2033..2040  
-10\_signal  
2059..2063  
/note="sigma H consensus"  
2253..2258  
RBS  
2270..3307  
CDS  
/note="pksofx2; similar to hamster  
hydroxymethylglutaryl-CoA synthase, Swiss-Prot Accession  
Number P13704"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AA085140.1"  
/db\_xref="GI:528991"  
/translation="NLITCSSESGIDFGKSLSTYIHEYLGILNRNCRLEFVKQACYSCTA  
GFQMAVNFILSQTSPGAKALVIASDISRFLIAEGDGLSEDSWYAEPSAGAVAVLV  
GENPEVFOIDPGANGYGVGVNDTCRPIPDSEAGSDLSMSYLDCCETQFLEYOKRV  
PGANYQDTFOYLAYHTPEGGVKGARHTMMKRVAKVKTSGIETDLTRYKPLNYCQR  
VGNMGALFLALASTIQGREDTPRIGCFSGYSGCCSEFYSGITTPQGGQRQRTFG  
IEKHLDRIYQLSMEEYELLFGSGMVRFGRNVKLDFFEMIPGIMOSTQEKPLFLLEEI  
SEFHRKYRWS"  
3285..3290  
RBS  
3295..4074  
CDS  
/note="pksofx3; similar to Rhodobacter capsulatus  
enoyl-CoA hydratase Swiss-Prot Accession Number P24162"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AA085141.1"  
/db\_xref="GI:528992"  
/translation="MDLYTYQTIKVRFOASVCYITFHRPEANNTINDTIEECLOVLN  
QCESTVTVVLEGLPEVFCFGADFOEYQEMKGRKQASQOEPLYDLMMKLTQGPYV  
TISHVRGNAGGLGFSATIDATADQATFASLSLELFLGYPACVLPFLIRIRGQKA  
HYMTLMTKPIISVQAESENGLIDAFDAESDVLRLHLLRLRLNKKGIAHYKQFMSSLD  
HOVSRKATALTANQDMFSDPQNGIIRYVETGQFPWEDQ"  
4103..4108  
RBS  
4114..4863  
CDS  
/note="pksofx4; similar to Rhodobacter capsulatus  
enoyl-CoA hydratase Swiss-Prot Accession Number P24162"  
/codon\_start=1  
/transl\_table=1  
/product="unknown"  
/protein\_id="AA085142.1"  
/db\_xref="GI:528993"





```

/rpt_family="AluJo"
complement(31059..31214)
/rpt_family="MER33"
31829..31896
/rpt_family="CAG)n"
31991..32280
/rpt_family="AluSg"
32281..32442
/rpt_family="AluSg/x"
32753..32794
/rpt_family="(GGA)n"
33138..33190
/rpt_family="(TG)n"
33285..33315
/rpt_family="(CCCCCG)n"
34350..34489
/rpt_family="AluJo"
34490..34784
/rpt_family="AluSx"
34785..34963
/rpt_family="AluJo"

```

Query Match 44.28; Score 23; DB 85; Length 120766;  
 Best Local Similarity 74.48; Pred. No. 1.8e+02;  
 Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

QY 1 tcgttcactcactcgagcggcgttaagcggccct 39
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 53624 TCGCGCGCTACTGGGGCGCGCTTCGCCCGCTGCT 53662

```

```

RESULT 10
BSUB0009 208780 bp DNA BCT 26-NOV-1997
LOCUS Bacillus subtilis complete genome (section 9 of 21): from 1598421
to 1807200.
ACCESSION Z99112 AL009126
VERSION Z99112.1 GI:2633902
KEYWORDS
SOURCE
ORGANISM Bacillus subtilis.
REFERENCE
AUTHORS Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Boursier, L., Brans, A., Braun, M., Brignell, S.,
Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E.,
Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R.,
Henaoui, A., Hilbert, H., Holsappel, S., Hosono, S., Hulio, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaer-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinis, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Masuda, S., Naul, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogilwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portetelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Fujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowska, A., Serot, S.J., Serror, P., Shin, B.S., Soldo, B.,
Sorokin, A., Taccini, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tanakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A.,
Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T.,
Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and

```

Danchin, A.  
 The complete genome sequence of the gram-positive bacterium  
 Bacillus subtilis  
 Nature 390 (6657), 249-256 (1997)  
 98044033  
 2 (bases 1 to 208780)  
 Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.  
 Direct Submission  
 Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,  
 Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724  
 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,  
 adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0) 45  
 68 89 48

FEATURES  
 source  
 1..208780  
 /organism="Bacillus subtilis"  
 /strain="168"  
 /db\_xref="taxon:1423"  
 258..4588  
 /gene="bpr"  
 258..4559  
 /gene="bpr"  
 /EC\_number="3.4.21.-"  
 /note="alternate gene name: bpf"  
 /codon\_start=1  
 /transl\_table=11  
 /product="bacillopeptidase F"  
 /protein\_id="CAB13404.1"  
 /db\_xref="GI:2633903"  
 /db\_xref="SWISS-PROT:P16397"  
 /translation="MRKTKNRLISSVLSTVTVSSLLFPGAAGASSKVTSPVSKKEIQ  
 SAESIQNKISSLKSKFKKERTFLIKFDLANPEKAAKAAVVKAKSKYLSAAKTEY  
 QKRSAYVSSLKTADESQQDLKYLNTQDKGNADQIHSYVYVNGIADVHASKVMEKV  
 VQPEVEKVLNPKRQLFKSSSPNNKKAOKAIKATDGVEMNVQDIDAPKAWALGYDG  
 TGVVVAIDTGVEMNHPALEKRYGYNPENPENEMMYDAVAGEASVDDLAHGT  
 HVTGTWVSGPDGTNIGVAPGAKNIAYKAFSEGGDDADILEAGEASVLPKDAEAGNP  
 HPEMAPDVYNNWGGGLDEWYRDYVNWRAADIFPEFSAGNTDIFPGPGSIANP  
 ANYPEFATGATDINKKLADFLQGGSPYDEIKPEISAPGVNIRSSVPGTYEDWDG  
 TSMAGPHVSAAALAKOANASLSDVEMEDILTSTAEPLTDTFTDPSPNNGYHGLVNA  
 FDVAVSVDGLGKAEGVSVGGDDPPVYHEKVTEAYEGSLPLTLTAEDNVSVTS  
 VKLSYKLDGEMTEITAKRISGDHLKGTQAEIPDLKGTLSYKMWIHDRGGHVSSD  
 VYDVTKPSITACYKODFETAPGWASGTNNWNGVSPSTGNTAASGEKVTGNTLT  
 GNYANSANNMLVMPPIKAPDSGLFLQKSWHNLEDDFDYGVFVLPGEKKNMEQAV  
 YNKTSTWTDDEEIDLSAYKGNIOVMFNQSDSIEIAKEGWYIDVVLSDSKAGTKVK  
 NLKGVPKSGKQKKKPNPKKAPKASANTAVKHONKAIQPOVLPLKQVSVVETGKSTY  
 SDQSTGYTLKHKAGDYTLMAEAYGQSKTKVSLKTDOTTOANFTLEEMKGTLKGT  
 VINKTTEPVTGASVYVEDAEPAMTNDKGYMLEAEGVATIKVAAPGYSDERS  
 VELKGDVTKETALKPVGYEGEYADGTAENANSYFAAGNHAYKMTLDGKDKGML  
 TGLGFPEWTEFPDGPTEKFEVYDATKDGAPKRIAGPFNAELRNGEWTKVDLS  
 SKGIWVDKDFLYIIOGKPPYSPGLAMDETQNSGRNNQYIDGKWPGDKADGNTMI  
 RALVDYEAAPVETSPDKSYTNKDSYTVKGNASPGTTHIYNGEKEAGETKAAADGT  
 FHAGIILNKGENELTASTDNGTDASSPITVTLQEKPELTLDNPKDGGKTKTKTV  
 TVKGAYSDNLKDVKNKATVADGSYSARILLENGRNEIKVIATDLNAGNKKTKTV  
 IDVNFQKPVISGLIPGEDKNLKAESVKIAFSAEDLDATFTIRMLPTNARASVQNA  
 ELPRLRISPGRYEGYWTATSSIKAKAKVEVIVRDDYDYGNETRKTANGKLNMTEN"  
 4571..4588  
 /gene="bpr"  
 4754..5683  
 /gene="spoilGA"  
 4754..5683  
 /gene="spoilGA"  
 /function="processing of pro-sigma-E (SpoIIGB) to active  
 sigma-E (stage II sporulation)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="protease"  
 /protein\_id="CAB13405.1"  
 /db\_xref="GI:2633904"  
 /db\_xref="SWISS-PROT:P13801"  
 /translation="MKIYLDVLIWLLNFCFDALLLLTAFILKRHHVKRRLVGVGAFTGS  
 SILVLMFTFSPVHPAGKLAFFVIVVYTFGKFRFFQNLFSFVATFLMGGI  
 IGASHLLQNSIVQNGVMTNQTGDPDISLFIIVGDFPALFFSKRRLEDIETKNIQ  
 YEERSVQADLGSQTLHVRGLIDSGNQLYDPLTKTPVMIIYIDKLEPIGTAEITMR

terminator

gene

CDS



```

NTDLEAIEQLDDSFRLDKMRLIPYRGVGOONQFLLCVKPDHVTIMTKEMISADK
5746..6465
/gene="sigE"
CDS
5746..6465
/gene="sigE"
/function="early mother cell-specific gene expression"
/notes="alternate gene name: spoIIGB"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma-29
factor (sigma-E)"
/protein_id="CAB13406.1"
/db_xref="GI:2633905"
/db_xref="SWISS-PROT:P06222"
/translation="MKLKLRLTHWYLLMKLGLKSDVYYIGSEALPPPLSKDEE
QVLLMKLPNGDQAARILIERILRVLYIARFENGINIEDLISIGTIGLKAYNTF
NPEKKIATYASRCILEMEILMYLRNKRIRSEVSYFPLNIDWDGNELLSDVLGTD
DDITTKDEANVDYDKLLKALKLEQLNEREKQIMELRFLVGBEEKTKQDVADMMGISQS
YISRLKRIIRKLRKEFNKV"
6605..7387
/gene="sigG"
CDS
6605..7387
/gene="sigG"
/function="late foresore-specific gene expression"
/notes="alternate gene name: spoIIIG"
/codon_start=1
/transl_table=11
/product="RNA polymerase sporulation-specific sigma factor
(sigma-G)"
/protein_id="CAB13407.1"
/db_xref="GI:2633906"
/db_xref="SWISS-PROT:P19940"
/translation="MSRNKVEICGVDTSKLPVLKNEMRKFLROLQDGGDSAREKLV
NGNRLVSVIQRNNGEYVDLQFQVCGICGLMKSIDNFDLSHNKVFSTYAVPMIIGE
LRLRLDNRPVRSLSLDIAYKALQVRERLISKTEPTAEKLVLEVPREEIVFA
LDAIDQVSLSEPIYNDGGDPIYHQDISDERNTDSQWIEELALKEGRMRRLNREKMI
LKRFPQKTQMEVAEEIGISOAQVSRLEKAAIKQMNKNH"
7535..8496
/gene="ylmA"
CDS
7535..8496
/gene="ylmA"
/function="unknown"
/notes="similar to ABC transporter (ATP-binding protein)"
/codon_start=1
/transl_table=11
/protein_id="CAB13408.1"
/db_xref="GI:2633907"
/db_xref="SPTREMBL:O31723"
/translation="MILQLDNVSLKRNCKWILKDIHWKVEKENWVLYGLNAGKTAL
LNMCSYFPTSGEMRVLGHEFGKTELGKLRKIGLVSAALQKLYPADSAFETALS
GAYASIGLYETPSKTEKATGLLEDLGAIEYADRYETLSQGEKQKRALIARALMADP
ELLIDEPVTGLDTIAREKLLDTITYIANKENAPSILYVTHAEIILPVFDKALLKQ
GEVFGSGIKEMLTQILSAFEDTPIHVLWNQDRPFLTRAEPITNA"
8472..8496
/gene="ylmB"
terminator
8531..9811
/gene="ylmB"
CDS
8531..9811
/gene="ylmB"
/function="unknown"
/notes="similar to acetylornithine deacetylase"
/codon_start=1
/transl_table=11
/protein_id="CAB13409.1"
/db_xref="GI:2633908"
/db_xref="SPTREMBL:O31724"
/translation="MDQIYSLQKKVEKKEELIQAKTLISVQTAPPARNTEGIOS
WIAGLYNELGFSIDKWDYPCDENVVGRKGTDSADYSLIINGHVDVDAEVEKEENK
HDPHPTEKNGLLIGRASDMKGMACVLFVKVLIRASIELPGDLDLIQSVIGEEVGE
AGTCCCKRGTHADFAIVADTSDMHIOGGVITGWIEIKSSQTFHDGTRRNHAGG
GTGASAEKMAKTIAGVADTSDMHIOGGVITGWIEIKSSQTFHDGTRRNHAGG
CRLWTVHFPNETHDQVAAEIEDYVNRLLSDSDIWLRENRPVFKWGGSSMIEDRGEIF

```

```

PALEVDPCHPGVLTALTASHQKVKRECPIDVDSQSVTDGGWLYDAGIPCVIYGPCLDHN
AHSVNEKVSIEQLVEYTKIILDPIISWCSRKKEQ"
9894..10139
/gene="ylmC"
CDS
9894..10139
/gene="ylmC"
/function="unknown"
/notes="similar to hypothetical proteins"
/codon_start=1
/transl_table=11
/protein_id="CAB13410.1"
/db_xref="GI:2633909"
/db_xref="SPTREMBL:O31725"
/translation="MISISEFQVKDVVNVSNKKLGSIGDIDINVTGKIQAIIILGNN
44.2%; Score 23; DB 2; Length 208780;
Query Match
Best Local Similarity 74.4%; Pred. No. 1.6e+02;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 13 ctgcgggacggcgtctaaagcggcccttaacacaaacgtt 51
|||||
Db 199854 CTGGGGGACCCGATCGATCGCTGGCGGCTTTAACAAAGCTT 199892

RESULT 11
AC023855 218864 bp DNA HTG 10-JAN-2001
LOCUS
AC023855
DEFINITION
Homo sapiens chromosome 17 clone RP11-546M21 map 17, WORKING DRAFT
SEQUENCE, 21 unordered pieces.
AC023855
ACCESSION
AC023855.5 GI:12061510
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (Lases 1 to 218864)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens chromosome 17, clone RP11-546M21
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 218864)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Goyette,M., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Larcocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meidrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Nayler,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M.,
Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,E.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (18-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 10, 2001 this sequence version replaced gi:8954143.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www-seq.wi.mit.edu

```



Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borris, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D., Emerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Golightly, E.J., Grandi, G., Guisepi, G., Guy, B.J., Haga, K., Haiech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Knaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidis, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Nael, C., Medigue, C., Medina, N., Mellado, R.P., Mizuno, M., Moesti, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serror, P., Shin, B.S., Soldo, B., Sorokin, A., Taccon, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Takakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenecker, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.

# TITLE The complete genome sequence of the gram-positive bacterium

JOURNAL Nature 390 (6657), 249-256 (1997)  
 MEDLINE 98044033  
 REFERENCE 2 (bases 1 to 233780)  
 AUTHORS Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr  
 Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

## FEATURES

source  
 1. .233780  
 /organism="Bacillus subtilis"  
 /strain="168"  
 /db\_xref="taxon:1423"  
 75 - 744  
 /gene="pksA"  
 75 - 692  
 /gene="pksA"  
 /function="regulation of the polyketide synthase operon (pks)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional regulator"  
 /protein\_id="CAB13591.1"  
 /db\_xref="GI:2634091"  
 /db\_xref="SPTREMBL:O34381"  
 /translation="MPKQIDHEKRRQIAETRVILRMEGASARNIAKEAGLSIG ALRHYFSTQDELLAFAMKLVQEKVTDRIKDIAVRDLDPKQVLQILLEMVPTNEETIR EMEYWFATYARHKDMEDASHDGI FSGMRNLFIAYLDESDDLKONAKDIEARLYA LVDGLAHMLDPVRVKNKIRKRVIMOVESICVEDTRETQKRHP"  
 709 - 744  
 /gene="pksA"  
 882 - 1559  
 /gene="pksB"  
 882 - 1559  
 /gene="pksB"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13592.1"  
 /db\_xref="GI:2634092"

## terminator

gene  
 CDS

gene  
 CDS

/db\_xref="SPTREMBL:O34769"  
 /translation="MNLTYKVHPKTRYGQGTNYCYTIEDIVSRSAIVVDPQSWELSKI TTILSEAEELKVALTHSHYDHNVLDPITKPNQAYMSKEIDYIQPRCNLISL DDHOTISIGNTRACALITPGTAGGCMYLFSESTIGDITVFTGCGICEDDGSAREM KGIOTIKSEVSPHVRYPVPGHSFGKSPGSHKIDLYQHNIYFQIDKKKFFVKFRTRKNO KGFIDFK"  
 1932 - 2798  
 /gene="pksC"  
 1932 - 2798  
 /gene="pksC"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13593.1"  
 /db\_xref="GI:2634093"  
 /db\_xref="SPTREMBL:O34825"  
 /translation="MITYVFGGSGQKMGSLDFEFKELTDQADEILGYSIKRCL ENPYSNLKNTQFTOPALYVYNALSYLKKIRDEEYKPDFVAGHSGLGEYNALFAAEAFDF ETGLQVYKRGELMSLISNGMAVGLNEEQVAKALKKEYHLHDVIANVAPYQIV I SGKDETEKASILFETVETVMVPLNVSAGAFHSRYNNKAKEEFEEFLHAFYFSPPSI PVISNVYAKPTYEFEMKQTLADQINHSVKWTTDSISYLMKKGMHEFEVGPNGVLTGLI HRIKDAEAMPR"  
 3302 - 4192  
 /gene="pksD"  
 3302 - 4192  
 /gene="pksD"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13594.1"  
 /db\_xref="GI:2634094"  
 /db\_xref="SPTREMBL:O34877"  
 /translation="MNEPLVFMFSGGQYHMGKELFKENTVFRQSMLEMDUATAARR IGTIVEEYHPGKRVSDPDSILFSPHAFIMFIEYSLYKVLDEGRGIIYDYLGSWYEN SCDKKMLAIDLPQLNDHPQLPFGNSELISINDSHFVIGSEEDHIRKIMEDLKEQ ILQCLLPVYAFHSLSDPAESAYAEFLRSKQKPSIPVSSLTGSLCHVMDNFEP NAVKPMFREAIRYLESQHTCKFIDLGPSOTLAAFYKQLIPGDSADRCSSIIITPFHQ ELKMLNVEYFRTPERKFR"  
 4189 - 6141  
 /gene="pksE"  
 4189 - 6141  
 /gene="pksE"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13595.1"  
 /db\_xref="GI:2634095"  
 /db\_xref="SPTREMBL:O34787"  
 /translation="MITYVFGGSGQKMGQGLFEQYQHLTDQADQILGYSIKELCT EKSVDYNTHTETOPALYVYNALSYLKRVEETGRKPDFAAGHSGLGEYNALMAAGAFDF ETGLRLVKRGELMGRITGGMAAVIGLSKEQVTAVLEEHRLYDIDVANENTPOQIV I SGPKTEKARAVFENTKVKLPHPNVSAGAFHSRYMNEAKQVFKOYIDSFQAPLAI PVISNVYAEYHODRLKDTLSEONDVTWKTDTSIRFLMGRGEMFEAEIGPCTVLTGLI HRIKNAEPLTYIPKKNPAISAHLEKORNVQAGITAESLGSAEFKQDYHLTYAYLAGG MYRGIAKEMVKLRSRAGMGMFFGTGSLGSLKEVEDAIIAIGELGQAGYGINLNVHNM KHTESEKMDILLRNQVSIVSEAFSLVTPVLYRAKGVKRNQNGDVICSNLIAK ISRPEVAESFLSPAPENHKLGENITWNEALLRCIPMADDICVADSGGHTDGG VAYSLSMPAMTSLRDEMKKYQKIRVGAAGGIGTPEAAMAAPMLGADFTLTSINO QTVCAATSDVKDLLOQMVDQDTAYAPAGDMFESGSKVQLKKGKGVFFPARANKLYEL QRVGSIIEIAKMLAQLEEKYKRSV"  
 6780 - 8030  
 /gene="pksF"  
 6780 - 8030  
 /gene="pksF"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13596.1"  
 /db\_xref="GI:2634096"  
 /db\_xref="SWISS-PROT:P40804"  
 /translation="MTKCNLPVVVGTGVTASIGCKEDFASLLSGRHAFDVMKRS GROKDSRFIGAETASLSYDRLSKKMLRKASFSRAALVLTITAEWEAEALDDADSSRI GLVVGSGNFQORENFVEYERYQDRSGFISPAYGLSFMDSDLCGICTQFGITGLAYTV

gene  
 CDS

/db\_xref="SPTREMBL:O34769"  
 /translation="MNLTYKVHPKTRYGQGTNYCYTIEDIVSRSAIVVDPQSWELSKI TTILSEAEELKVALTHSHYDHNVLDPITKPNQAYMSKEIDYIQPRCNLISL DDHOTISIGNTRACALITPGTAGGCMYLFSESTIGDITVFTGCGICEDDGSAREM KGIOTIKSEVSPHVRYPVPGHSFGKSPGSHKIDLYQHNIYFQIDKKKFFVKFRTRKNO KGFIDFK"  
 1932 - 2798  
 /gene="pksC"  
 1932 - 2798  
 /gene="pksC"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13593.1"  
 /db\_xref="GI:2634093"  
 /db\_xref="SPTREMBL:O34825"  
 /translation="MITYVFGGSGQKMGSLDFEFKELTDQADEILGYSIKRCL ENPYSNLKNTQFTOPALYVYNALSYLKKIRDEEYKPDFVAGHSGLGEYNALFAAEAFDF ETGLQVYKRGELMSLISNGMAVGLNEEQVAKALKKEYHLHDVIANVAPYQIV I SGKDETEKASILFETVETVMVPLNVSAGAFHSRYNNKAKEEFEEFLHAFYFSPPSI PVISNVYAKPTYEFEMKQTLADQINHSVKWTTDSISYLMKKGMHEFEVGPNGVLTGLI HRIKDAEAMPR"  
 3302 - 4192  
 /gene="pksD"  
 3302 - 4192  
 /gene="pksD"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13594.1"  
 /db\_xref="GI:2634094"  
 /db\_xref="SPTREMBL:O34877"  
 /translation="MNEPLVFMFSGGQYHMGKELFKENTVFRQSMLEMDUATAARR IGTIVEEYHPGKRVSDPDSILFSPHAFIMFIEYSLYKVLDEGRGIIYDYLGSWYEN SCDKKMLAIDLPQLNDHPQLPFGNSELISINDSHFVIGSEEDHIRKIMEDLKEQ ILQCLLPVYAFHSLSDPAESAYAEFLRSKQKPSIPVSSLTGSLCHVMDNFEP NAVKPMFREAIRYLESQHTCKFIDLGPSOTLAAFYKQLIPGDSADRCSSIIITPFHQ ELKMLNVEYFRTPERKFR"  
 4189 - 6141  
 /gene="pksE"  
 4189 - 6141  
 /gene="pksE"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13595.1"  
 /db\_xref="GI:2634095"  
 /db\_xref="SPTREMBL:O34787"  
 /translation="MITYVFGGSGQKMGQGLFEQYQHLTDQADQILGYSIKELCT EKSVDYNTHTETOPALYVYNALSYLKRVEETGRKPDFAAGHSGLGEYNALMAAGAFDF ETGLRLVKRGELMGRITGGMAAVIGLSKEQVTAVLEEHRLYDIDVANENTPOQIV I SGPKTEKARAVFENTKVKLPHPNVSAGAFHSRYMNEAKQVFKOYIDSFQAPLAI PVISNVYAEYHODRLKDTLSEONDVTWKTDTSIRFLMGRGEMFEAEIGPCTVLTGLI HRIKNAEPLTYIPKKNPAISAHLEKORNVQAGITAESLGSAEFKQDYHLTYAYLAGG MYRGIAKEMVKLRSRAGMGMFFGTGSLGSLKEVEDAIIAIGELGQAGYGINLNVHNM KHTESEKMDILLRNQVSIVSEAFSLVTPVLYRAKGVKRNQNGDVICSNLIAK ISRPEVAESFLSPAPENHKLGENITWNEALLRCIPMADDICVADSGGHTDGG VAYSLSMPAMTSLRDEMKKYQKIRVGAAGGIGTPEAAMAAPMLGADFTLTSINO QTVCAATSDVKDLLOQMVDQDTAYAPAGDMFESGSKVQLKKGKGVFFPARANKLYEL QRVGSIIEIAKMLAQLEEKYKRSV"  
 6780 - 8030  
 /gene="pksF"  
 6780 - 8030  
 /gene="pksF"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13596.1"  
 /db\_xref="GI:2634096"  
 /db\_xref="SWISS-PROT:P40804"  
 /translation="MTKCNLPVVVGTGVTASIGCKEDFASLLSGRHAFDVMKRS GROKDSRFIGAETASLSYDRLSKKMLRKASFSRAALVLTITAEWEAEALDDADSSRI GLVVGSGNFQORENFVEYERYQDRSGFISPAYGLSFMDSDLCGICTQFGITGLAYTV

gene  
 CDS

/db\_xref="SPTREMBL:O34769"  
 /translation="MNLTYKVHPKTRYGQGTNYCYTIEDIVSRSAIVVDPQSWELSKI TTILSEAEELKVALTHSHYDHNVLDPITKPNQAYMSKEIDYIQPRCNLISL DDHOTISIGNTRACALITPGTAGGCMYLFSESTIGDITVFTGCGICEDDGSAREM KGIOTIKSEVSPHVRYPVPGHSFGKSPGSHKIDLYQHNIYFQIDKKKFFVKFRTRKNO KGFIDFK"  
 1932 - 2798  
 /gene="pksC"  
 1932 - 2798  
 /gene="pksC"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13593.1"  
 /db\_xref="GI:2634093"  
 /db\_xref="SPTREMBL:O34825"  
 /translation="MITYVFGGSGQKMGSLDFEFKELTDQADEILGYSIKRCL ENPYSNLKNTQFTOPALYVYNALSYLKKIRDEEYKPDFVAGHSGLGEYNALFAAEAFDF ETGLQVYKRGELMSLISNGMAVGLNEEQVAKALKKEYHLHDVIANVAPYQIV I SGKDETEKASILFETVETVMVPLNVSAGAFHSRYNNKAKEEFEEFLHAFYFSPPSI PVISNVYAKPTYEFEMKQTLADQINHSVKWTTDSISYLMKKGMHEFEVGPNGVLTGLI HRIKDAEAMPR"  
 3302 - 4192  
 /gene="pksD"  
 3302 - 4192  
 /gene="pksD"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13594.1"  
 /db\_xref="GI:2634094"  
 /db\_xref="SPTREMBL:O34877"  
 /translation="MNEPLVFMFSGGQYHMGKELFKENTVFRQSMLEMDUATAARR IGTIVEEYHPGKRVSDPDSILFSPHAFIMFIEYSLYKVLDEGRGIIYDYLGSWYEN SCDKKMLAIDLPQLNDHPQLPFGNSELISINDSHFVIGSEEDHIRKIMEDLKEQ ILQCLLPVYAFHSLSDPAESAYAEFLRSKQKPSIPVSSLTGSLCHVMDNFEP NAVKPMFREAIRYLESQHTCKFIDLGPSOTLAAFYKQLIPGDSADRCSSIIITPFHQ ELKMLNVEYFRTPERKFR"  
 4189 - 6141  
 /gene="pksE"  
 4189 - 6141  
 /gene="pksE"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13595.1"  
 /db\_xref="GI:2634095"  
 /db\_xref="SPTREMBL:O34787"  
 /translation="MITYVFGGSGQKMGQGLFEQYQHLTDQADQILGYSIKELCT EKSVDYNTHTETOPALYVYNALSYLKRVEETGRKPDFAAGHSGLGEYNALMAAGAFDF ETGLRLVKRGELMGRITGGMAAVIGLSKEQVTAVLEEHRLYDIDVANENTPOQIV I SGPKTEKARAVFENTKVKLPHPNVSAGAFHSRYMNEAKQVFKOYIDSFQAPLAI PVISNVYAEYHODRLKDTLSEONDVTWKTDTSIRFLMGRGEMFEAEIGPCTVLTGLI HRIKNAEPLTYIPKKNPAISAHLEKORNVQAGITAESLGSAEFKQDYHLTYAYLAGG MYRGIAKEMVKLRSRAGMGMFFGTGSLGSLKEVEDAIIAIGELGQAGYGINLNVHNM KHTESEKMDILLRNQVSIVSEAFSLVTPVLYRAKGVKRNQNGDVICSNLIAK ISRPEVAESFLSPAPENHKLGENITWNEALLRCIPMADDICVADSGGHTDGG VAYSLSMPAMTSLRDEMKKYQKIRVGAAGGIGTPEAAMAAPMLGADFTLTSINO QTVCAATSDVKDLLOQMVDQDTAYAPAGDMFESGSKVQLKKGKGVFFPARANKLYEL QRVGSIIEIAKMLAQLEEKYKRSV"  
 6780 - 8030  
 /gene="pksF"  
 6780 - 8030  
 /gene="pksF"  
 /function="involved in polyketide synthesis"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAB13596.1"  
 /db\_xref="GI:2634096"  
 /db\_xref="SWISS-PROT:P40804"  
 /translation="MTKCNLPVVVGTGVTASIGCKEDFASLLSGRHAFDVMKRS GROKDSRFIGAETASLSYDRLSKKMLRKASFSRAALVLTITAEWEAEALDDADSSRI GLVVGSGNFQORENFVEYERYQDRSGFISPAYGLSFMDSDLCGICTQFGITGLAYTV



```

VERSION      AF033487.1  GI:2668727
KEYWORDS
SOURCE       Eupenicillium tularensis.
ORGANISM     Eupenicillium tularensis.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eutiales; Trichosporaceae; Eupenicillium.
REFERENCE    1 (bases 1 to 1142)
AUTHORS      Peterson,S.W.
TITLE        Phylogenetic analysis of Penicillium species based on ITS and
JOURNAL      LSU-rDNA nucleotide sequences
              (in) Samson,R.A. and Pitt,J.I. (Eds.);
              INTEGRATION OF MODERN TAXONOMIC METHODS FOR PENICILLIUM AND
              ASPERGILLUS CLASSIFICATION: 163-178;
              HARWOOD Academic Publishers, The Netherlands (2000)
REFERENCE    2 (bases 1 to 1142)
AUTHORS      Peterson,S.W.
TITLE        Phylogenetic analysis of Penicillium species based on ITS and
JOURNAL      LSU-rDNA nucleotide sequences
REFERENCE    3 (bases 1 to 1142)
AUTHORS      Peterson,S.W.
TITLE        Direct Submission
JOURNAL      Submitted (07-NOV-1997) Microbial Properties Research, National
              Center for Agricultural Utilization Research, 1815 N. University
              St., Peoria, IL 61604, USA
FEATURES
source       Location/Qualifiers
              1..1142
                /organism="Eupenicillium tularensis"
                /strain="NRRL 5273"
                /db_xref="taxon:36626"
misc_RNA     1..184
              /product="internal transcribed spacer 1"
rRNA         185..341
              /product="5.8S"
misc_RNA     342..506
              /product="internal transcribed spacer 2"
rRNA         507..>1142
              /product="28S"
BASE COUNT   255 a 306 c 347 g 234 t
ORIGIN
Query Match 43.5%; Score 22.6; DB 12; Length 1142;
Best Local Similarity 68.9%; Pred. No. 6.9e+02;
Matches 31; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 7 cgcctactgcggcggtacgaagcggcccttaaccacagctt 51
   | | | | | | | | | | | | | | | | | | | | | |
Db 454 CCCGCTCGAGCGCGCGCGCGCGCGCGCGGATCAACCAACTTT 498
               | | | | | | | | | | | | | | | | | | | | | |

RESULT 15
AC084177     45257 bp  DNA  HTG  14-OCT-2000
LOCUS        Homo sapiens chromosome 16 clone RP11-616M22 map 16, LOW-PASS
DEFINITION   SEQUENCE SAMPLING.
AC084177
VERSION      AC084177.1  GI:10801405
KEYWORDS     HTG; HTGS_PHASE0.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 45257)
AUTHORS      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE        Homo sapiens chromosome 16, clone RP11-616M22
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 45257)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Barna,N., Bastien,V., Bedalov,F., Boguslavskiy,L.,
              Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
              Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
              DeAureliano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,

```

```

Fitzhugh,W., Cagle,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iltiev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (14-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11412
Center clone name: 616_M_22
-----
* NOTE: This record contains 55 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 756: contig of 756 bp in length
* 757 856: gap of 100 bp
* 857 1616: contig of 760 bp in length
* 1617 1716: gap of 100 bp
* 1717 2434: contig of 718 bp in length
* 2435 2534: gap of 100 bp
* 2535 3271: contig of 737 bp in length
* 3272 3371: gap of 100 bp
* 3372 4120: contig of 749 bp in length
* 4121 4220: gap of 100 bp
* 4221 4948: contig of 728 bp in length
* 4949 5048: gap of 100 bp
* 5049 5765: contig of 717 bp in length
* 5766 5865: gap of 100 bp
* 5866 6564: contig of 699 bp in length
* 6565 6664: gap of 100 bp
* 6665 7376: contig of 712 bp in length
* 7377 7476: gap of 100 bp
* 7477 8197: contig of 721 bp in length
* 8198 8297: gap of 100 bp
* 8298 9048: contig of 751 bp in length
* 9049 9148: gap of 100 bp
* 9149 9872: contig of 724 bp in length
* 9873 9972: gap of 100 bp
* 9973 10697: contig of 725 bp in length
* 10698 10797: gap of 100 bp
* 10798 11525: contig of 728 bp in length
* 11526 11625: gap of 100 bp
* 11626 12347: contig of 722 bp in length
* 12348 12447: gap of 100 bp
* 12448 13188: contig of 741 bp in length
* 13189 13288: gap of 100 bp

```

TITLE  
JOURNAL  
COMMENT

*	13289	14038:	contig of 750 bp	in length
*	14039	14138:	gap of 100 bp	
*	14139	14862:	contig of 724 bp	in length
*	14863	14962:	gap of 100 bp	
*	14963	15677:	contig of 715 bp	in length
*	15678	15777:	gap of 100 bp	
*	15778	16491:	contig of 714 bp	in length
*	16492	16591:	gap of 100 bp	
*	16592	17342:	contig of 751 bp	in length
*	17343	17442:	gap of 100 bp	
*	17443	18158:	contig of 716 bp	in length
*	18159	18258:	gap of 100 bp	
*	18259	18954:	contig of 696 bp	in length
*	18955	19054:	gap of 100 bp	
*	19055	19784:	contig of 730 bp	in length
*	19785	19884:	gap of 100 bp	
*	19885	20594:	contig of 710 bp	in length
*	20595	20694:	gap of 100 bp	
*	20695	21423:	contig of 729 bp	in length
*	21424	21523:	gap of 100 bp	
*	21524	22231:	contig of 708 bp	in length
*	22232	22331:	gap of 100 bp	
*	22332	23044:	contig of 713 bp	in length
*	23045	23144:	gap of 100 bp	
*	23145	23867:	contig of 723 bp	in length
*	23868	23967:	gap of 100 bp	
*	23968	24678:	contig of 711 bp	in length
*	24679	24778:	gap of 100 bp	
*	24779	25498:	contig of 720 bp	in length
*	25499	25598:	gap of 100 bp	
*	25599	26317:	contig of 719 bp	in length
*	26318	26417:	gap of 100 bp	
*	26418	27169:	contig of 752 bp	in length
*	27170	27269:	gap of 100 bp	
*	27270	28003:	contig of 734 bp	in length
*	28004	28103:	gap of 100 bp	
*	28104	28826:	contig of 723 bp	in length
*	28827	28926:	gap of 100 bp	
*	28927	29620:	contig of 694 bp	in length
*	29621	29720:	gap of 100 bp	
*	29721	30444:	contig of 724 bp	in length
*	30445	30544:	gap of 100 bp	
*	30545	31250:	contig of 706 bp	in length
*	31251	31350:	gap of 100 bp	
*	31351	32081:	contig of 731 bp	in length
*	32082	32181:	gap of 100 bp	
*	32182	32915:	contig of 734 bp	in length
*	32916	33015:	gap of 100 bp	
*	33016	33719:	contig of 704 bp	in length
*	33720	33819:	gap of 100 bp	
*	33820	34537:	contig of 718 bp	in length
*	34538	34637:	gap of 100 bp	
*	34638	35330:	contig of 693 bp	in length
*	35331	35430:	gap of 100 bp	
*	35431	36154:	contig of 724 bp	in length
*	36155	36254:	gap of 100 bp	
*	36255	36975:	contig of 721 bp	in length
*	36976	37075:	gap of 100 bp	
*	37076	37807:	contig of 732 bp	in length
*	37808	37907:	gap of 100 bp	
*	37908	38645:	contig of 738 bp	in length
*	38646	38745:	gap of 100 bp	
*	38746	39463:	contig of 718 bp	in length
*	39464	39563:	gap of 100 bp	
*	39564	40294:	contig of 731 bp	in length
*	40295	40394:	gap of 100 bp	
*	40395	41115:	contig of 721 bp	in length
*	41116	41215:	gap of 100 bp	
*	41216	41935:	contig of 720 bp	in length
*	41936	42035:	gap of 100 bp	
*	42036	42775:	contig of 740 bp	in length
*	42776	42875:	gap of 100 bp	
*	42876	43620:	contig of 745 bp	in length

```

* 43621 43720: gap of 100 bp
* 43721 44444: contig of 724 bp in length
* 44445 44544: gap of 100 bp
* 44545 45257: contig of 713 bp in length.

FEATURES
    source
        1. 45257
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /map="16"
            /clone="RP11-616W22"
            /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT      7145 a 12298 c 10826 g 8936 t 6052 others
ORIGIN

Query Match      43.5%; Score 22.6; DB 76; Length
Best Local Similarity 67.4%; Pred. No. 3,1e+02;
Matches 31; Conservative 0; Mismatches 15; Indels

Qy 4 cttcgctcaatcgggaccgagcgtctaaagccggcccttaaccacaacg 49
      ||||| ||| ||||| | | ||||| ||| ||
Db 39102 CCTGCCTCCATCGNGGACCGGGTGGGAGGCCCGCTTACCGGTGCG 147

```

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw 1.0

Run on: August 2, 2001, 16:55:10 ; Search time 1874.14 Seconds  
(without alignments)  
388.375 Million cell updates/sec

Title: US-09-701-626A-78  
Perfect score: 77  
Sequence: 1 ccttaacaaagtgttcaagc.....cccttaacccaacgttagag 77

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_est4:*
5:	gb_est5:*
6:	gb_est6:*
7:	gb_est7:*
8:	gb_est8:*
9:	gb_est9:*
10:	gb_est10:*
11:	gb_est11:*
12:	gb_est12:*
13:	gb_est13:*
14:	gb_est14:*
15:	gb_est15:*
16:	gb_est16:*
17:	gb_est17:*
18:	gb_est18:*
19:	gb_est19:*
20:	gb_est20:*
21:	gb_est21:*
22:	gb_est22:*
23:	gb_est23:*
24:	gb_est24:*
25:	gb_est33:*
26:	gb_est34:*
27:	gb_est35:*
28:	gb_est36:*
29:	gb_est37:*
30:	gb_est38:*
31:	gb_est39:*
32:	gb_est40:*
33:	em_estba:*
34:	em_estfun:*
35:	em_esthum1:*
36:	em_esthum2:*
37:	em_esthum3:*
38:	em_esthum4:*
39:	em_esthum5:*
40:	em_esthum6:*
41:	em_esthum7:*
42:	em_esthum8:*
43:	em_esthum9:*
44:	em_esthum10:*
45:	em_esthum11:*
46:	em_esthum12:*
47:	em_esthum13:*
48:	em_esthum14:*
49:	em_esthum15:*
50:	em_esthum16:*
51:	em_esthum17:*
52:	em_esthum18:*
53:	em_esthum19:*
54:	em_esthum20:*
55:	em_esthum21:*
56:	em_esthum22:*
57:	em_esthum23:*
58:	em_esthum24:*
59:	em_esthum25:*
60:	em_esthum26:*
61:	em_esthum27:*
62:	em_esthum28:*
63:	em_estin1:*
64:	em_estin2:*
65:	em_estin3:*
66:	em_estin4:*
67:	em_estin5:*
68:	em_estom1:*
69:	em_estom2:*
70:	em_estov1:*
71:	em_estov2:*
72:	em_estpl1:*
73:	em_estpl2:*
74:	em_estpl3:*
75:	em_estpl4:*
76:	em_estpl5:*
77:	em_estpl6:*
78:	em_estpl7:*
79:	em_estpl8:*
80:	em_estpl9:*
81:	em_estpl10:*
82:	em_estro1:*
83:	em_estro2:*
84:	em_estro3:*
85:	em_estro4:*
86:	em_estro5:*
87:	em_estro6:*
88:	em_estro7:*
89:	em_estro8:*
90:	em_estro9:*
91:	em_estro10:*
92:	em_estro11:*
93:	em_estro12:*
94:	em_estro13:*
95:	em_estro14:*
96:	em_estro15:*
97:	em_estro16:*
98:	em_estro17:*
99:	em_estro18:*
100:	em_estro19:*
101:	em_estro20:*
102:	gb_est25:*
103:	gb_est26:*
104:	gb_est27:*
105:	gb_est28:*
106:	gb_est29:*
107:	gb_est30:*
108:	gb_est31:*
109:	gb_est32:*
110:	gb_est41:*
111:	gb_est42:*
112:	gb_est43:*
113:	gb_est44:*
114:	gb_est45:*
115:	gb_est46:*
116:	gb_est47:*



117: gb\_est48:\*  
118: gb\_est49:\*  
119: gb\_est50:\*  
120: gb\_est51:\*  
121: gb\_est52:\*  
122: gb\_est53:\*  
123: gb\_est54:\*  
124: gb\_est55:\*  
125: gb\_est56:\*  
126: gb\_est57:\*  
127: gb\_est58:\*  
128: gb\_est59:\*  
129: gb\_est60:\*  
130: gb\_est61:\*  
131: gb\_est62:\*  
132: gb\_est63:\*  
133: gb\_est64:\*  
134: gb\_est65:\*  
135: gb\_est66:\*  
136: gb\_est67:\*  
137: gb\_est68:\*  
138: gb\_est69:\*  
139: gb\_est70:\*  
140: gb\_est71:\*  
141: gb\_est72:\*  
142: gb\_est73:\*  
143: gb\_est74:\*  
144: gb\_est75:\*  
145: gb\_est76:\*  
146: gb\_est77:\*  
147: gb\_est78:\*  
148: gb\_est79:\*  
149: gb\_est80:\*  
150: gb\_est81:\*  
151: gb\_est82:\*  
152: gb\_est83:\*  
153: gb\_est84:\*  
154: gb\_est85:\*  
155: gb\_est86:\*  
156: gb\_est87:\*  
157: gb\_est88:\*  
158: gb\_est89:\*  
159: gb\_est90:\*  
160: gb\_est91:\*  
161: gb\_est92:\*  
162: gb\_est93:\*  
163: gb\_est94:\*  
164: gb\_est95:\*  
165: gb\_est96:\*  
166: gb\_est97:\*  
167: gb\_est98:\*  
168: gb\_est99:\*  
169: gb\_est100:\*  
170: gb\_est101:\*  
171: gb\_est102:\*  
172: gb\_est103:\*  
173: gb\_est104:\*  
174: gb\_est105:\*  
175: gb\_est106:\*  
176: gb\_est107:\*  
177: gb\_est108:\*  
178: gb\_est109:\*  
179: gb\_est110:\*  
180: gb\_est111:\*  
181: gb\_est112:\*  
182: gb\_est113:\*  
183: gb\_est114:\*  
184: gb\_est115:\*  
185: gb\_est116:\*  
186: gb\_est117:\*  
187: gb\_est118:\*  
188: gb\_est119:\*  
189: gb\_est120:\*

190: gb\_est121:\*  
191: gb\_est122:\*  
192: gb\_est123:\*  
193: gb\_est124:\*  
194: gb\_est125:\*  
195: gb\_est126:\*  
196: gb\_est127:\*  
197: gb\_est128:\*  
198: gb\_est129:\*  
199: gb\_est130:\*  
200: gb\_est131:\*  
201: gb\_est132:\*  
202: gb\_est133:\*  
203: gb\_est134:\*  
204: gb\_est135:\*  
205: gb\_est136:\*  
206: gb\_est137:\*  
207: gb\_est138:\*  
208: gb\_est139:\*  
209: gb\_est140:\*  
210: gb\_est141:\*  
211: gb\_est142:\*  
212: gb\_est143:\*  
213: gb\_est144:\*  
214: gb\_est145:\*  
215: gb\_est146:\*  
216: gb\_est147:\*  
217: gb\_est148:\*  
218: gb\_est149:\*  
219: gb\_est150:\*  
220: gb\_est151:\*  
221: gb\_est152:\*  
222: gb\_est153:\*  
223: gb\_est154:\*  
224: gb\_est155:\*  
225: gb\_est156:\*  
226: gb\_est157:\*  
227: gb\_est158:\*  
228: gb\_est159:\*  
229: gb\_est160:\*  
230: gb\_est161:\*  
231: gb\_est162:\*  
232: gb\_est163:\*  
233: gb\_est164:\*  
234: gb\_est165:\*  
235: gb\_est166:\*  
236: gb\_est167:\*  
237: gb\_est168:\*  
238: gb\_est169:\*  
239: gb\_est170:\*  
240: gb\_est171:\*  
241: gb\_est172:\*  
242: gb\_est173:\*  
243: gb\_est174:\*  
244: gb\_est175:\*  
245: gb\_est176:\*  
246: gb\_est177:\*  
247: gb\_est178:\*  
248: gb\_est179:\*  
249: gb\_est180:\*  
250: gb\_est181:\*  
251: gb\_est182:\*  
252: gb\_est183:\*  
253: gb\_est184:\*  
254: gb\_est185:\*  
255: gb\_est186:\*  
256: gb\_est187:\*  
257: gb\_est188:\*  
258: gb\_est189:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



```

/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT      86 a 186 c 204 g 134 t
ORIGIN

Query Match      35.8%; Score 27.6; DB 138; Length 610;
Best Local Similarity 63.6%; Pred. No. 8;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 9 aatggtcaagccgttccttcgctcactcggaccgactaaagccgccccttaaccaa 68
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 84 AAGATTCAAGCCTTCCTAGTCTCCGTATGCGCCGCCCAACGCCCGCTCTCGGCAT 25

Qy 69 acgtta 74
||| ||
Db 24 ACGGTA 19

RESULT 3
AL524576      536 bp mRNA EST 13-FEB-2001
LOCUS AL524576 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC007YG04 5
DEFINITION prime, mRNA sequence.
ACCESSION AL524576
VERSION AL524576.1 GI:12788069
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 536)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. 536
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC007YG04"
/clone_lib="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies, Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      96 a 172 c 178 g 74 t 16 others
ORIGIN

Query Match      35.1%; Score 27; DB 105; Length 536;
Best Local Similarity 57.5%; Pred. No. 13;
Matches 42; Conservative 3; Mismatches 28; Indels 0; Gaps 0;

Qy 3 ctaacaatggttcaagccgttccttcgctcactcggaccgactaaagccgcccctt 62
||||| : ||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 241 CTAMCARCTGGTGACGGCGCGCTTCAMGCTCTGGGGCCGAGCGCGGCGCTG 300

```

```

Qy 63 aaccaaagcttag 75
||||| |||||
Db 301 GACCAAGTCGCGAG 313

RESULT 4
AA203709      675 bp mRNA EST 24-JAN-1997
LOCUS AA203709
DEFINITION zx52d10.r1 Soares_fetal_liver_spleen_lNFLS_S1 Homo sapiens cDNA
clone IMAGE:446131 5', similar to contains element OFR repetitive
element ; mRNA sequence.
ACCESSION AA203709
VERSION AA203709.1 GI:1799436
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 675)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohldmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 492.
FEATURES
source
1. 675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:446131"
/clone_lib="Soares_fetal_liver_spleen_lNFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen lNFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AACTGGAAGAAATAATTAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      191 a 159 c 138 g 185 t 2 others
ORIGIN

Query Match      33.8%; Score 26; DB 3; Length 675;
Best Local Similarity 67.3%; Pred. No. 31;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 11 tggttcaagccgttcgcttcgctcactcggaccggttaagccgcccctt 62
|| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 562 TGGTTCAAGCCTGTGCTCTGCTGCTCACTCCCGACCGCTGGAGCCNTACNTT 613

RESULT 5
AQ869312      757 bp DNA GSS 03-NOV-1999
LOCUS AQ869312
DEFINITION nbcb0034P16r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic

```



```

REFERENCE
AUTHORS      1 (bases 1 to 905)
              Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
              Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
              Weissenbach,J.
TITLE        Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetraodon nigroviridis
JOURNAL      Unpublished
AUTHORS      2 (bases 1 to 905)
              Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
              Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
              Saurin,W. and Weissenbach,J.
TITLE        Human gene number estimate provided by genome wide analysis using
              Tetraodon nigroviridis DNA sequence
JOURNAL      Unpublished
AUTHORS      3 (bases 1 to 905)
              Direct Submission
TITLE        Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL      This sequence is a single read and was generated as part of a large
COMMENT      scale clone-end sequencing project of the Tetraodon nigroviridis
              genome. For more information, please take a look at
              http://www.genoscope.cns.fr/tetraodon.
FEATURES     Location/Qualifiers
              source
                1..905
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="007K24"
                /note="Genoscope sequence ID : COBH007BF12XD1-end : 77"
BASE COUNT   233 a 305 c 207 g 160 t
ORIGIN
Query Match 33.5%; Score 25.8; DB 221; Length 905;
Best Local Similarity 63.9%; Pred. No. 38;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 ctctacaaatggttcacgcttcgctcactcggtacggcgctaaagcgccccc 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 CTCACCAAGAAATGATCCCTCGCTCGTCCCTCAGTCCGGAGCAGCTCAAGCAGCACCC 434

Qy 61 t 61
Db 435 T 435

RESULT 8
CNS05S0L    1092 bp    DNA    GSS    26-MAY-2000
LOCUS       Tetraodon nigroviridis genome survey sequence T3 end of clone
DEFINITION  07C15 of library A from Tetraodon nigroviridis, genomic survey
              sequence.
ACCESSION   AL351246
VERSION     AL351246.1 GI:8245016
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM    Tetraodon nigroviridis
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
              Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 1092)
AUTHORS     Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
              Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
              Weissenbach,J.
TITLE       Characterization and repeat analysis of the compact genome of the
              freshwater pufferfish Tetraodon nigroviridis
JOURNAL     Unpublished
AUTHORS     2 (bases 1 to 1092)
              Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
              Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
              Saurin,W. and Weissenbach,J.
TITLE       Human gene number estimate provided by genome wide analysis using

```

```

Tetraodon nigroviridis DNA sequence
Unpublished
REFERENCE    3 (bases 1 to 1092)
AUTHORS      Direct Submission
TITLE        Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL      This sequence is a single read and was generated as part of a large
COMMENT      scale clone-end sequencing project of the Tetraodon nigroviridis
              genome. For more information, please take a look at
              http://www.genoscope.cns.fr/Tetraodon.
FEATURES     Location/Qualifiers
              source
                1..1092
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="007C15"
                /clone.lib="A"
                /note="Genoscope sequence ID : COAA007AB08A1-end : T3"
BASE COUNT   263 a 367 c 240 g 207 t 15 others
ORIGIN
Query Match 33.5%; Score 25.8; DB 222; Length 1092;
Best Local Similarity 63.9%; Pred. No. 39;
Matches 39; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1 ctctacaaatggttcacgcttcgctcactcggtacggcgctaaagcgccccc 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 820 CTCCAACGAAATGATCCCTCGCTCGTCCCTCAGTCCGGAGCAGCTCAAGCAGCACCC 879

Qy 61 t 61
Db 880 T 880

RESULT 9
AW010221    674 bp    mRNA    EST    10-SEP-1999
LOCUS       ST03E05 Pine Triplex shoot tip library Pinus taeda cDNA clone
DEFINITION  ST03E05, mRNA sequence.
ACCESSION   AW010221
VERSION     AW010221.1 GI:5858999
KEYWORDS    EST.
SOURCE      loblolly pine.
ORGANISM    Pinus taeda
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE   1 (bases 1 to 674)
AUTHORS     Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
TITLE       The Pine Gene Discovery Project
JOURNAL     Unpublished (1999)
COMMENT     Contact: Ross Whetten
              Forest Biotechnology Group
              North Carolina State University
              Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
              ,NC, 27695-8008
              Tel: 919-515-7800
              Fax: 919-515-7801
              Email: rosswhetten@unity.ncsu.edu
              Seq primer: 5' lambda Triplex2 Sequencing Primer.
FEATURES     Location/Qualifiers
              source
                1..674
                /organism="Pinus taeda"
                /db_xref="taxon:3352"
                /clone="ST03E05"
                /clone.lib="Pine Triplex shoot tip library"
                /note="Organ: shoot tips; Vector: Lambda Triplex; Site:1;
                Sfil (A); Site:2: Sfil (B); Shoot tips (approx. 2 cm from
                apex) were collected during the spring, frozen and used
                for mRNA isolation. The SMART-PCR method (Clontech) was
                used to prepare a library from 1 ug total RNA, using the
                Lambda Triplex vector. Plasmid subclones in pTriplex were
                recovered by cre-lox excision in E. coli strain BM25.8 and

```

```

BASE COUNT      179 a      158 c      195 g      129 t      13 others
ORIGIN

Query Match      33.2%; Score 25.6; DB 110; Length 674;
Best Local Similarity 58.1%; Pred. No. 44;
Matches 43; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 4 taacaaatggttcacgcttcgtctcactcgcggaccgctaaagcggccctta 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 447 TAGCAATTGATGAAGCGCAAGCATCCCTTATNACAAAATGCTAGCCCTCGCCTCC 506
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 64 accaaacgttag 77
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 507 CCCAAGNTTAG 520

RESULT 10
AU001605      532 bp      mRNA      EST      15-JAN-1999
LOCUS
DEFINITION AU001605 Bombyx mori p50(Daizo) Bombyx mori cDNA clone fbm1062f,
mRNA sequence.
ACCESSION AU001605
VERSION AU001605.1 GI:4157849
KEYWORDS EST.
SOURCE domestic silkworm.
ORGANISM Bombyx mori
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
; Bombycoidea; Bombycidae; Bombyx.
REFERENCE 1 (bases 1 to 532)
AUTHORS Mita, K., Morimyo, M., Shimada, T., Okano, K. and Maeda, S.
TITLE Establishment of cDNA database of Bombyx mori
JOURNAL Unpublished (1999)
COMMENT Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: knita@nirs.go.jp
PROJECT = 'CREST project by JST'.
FEATURES
    source
        1..532
            /organism="Bombyx mori"
            /strain="p50(Daizo)"
            /db_xref="taxon:7091"
            /clone="fbm1062f"
            /clone_lib="Bombyx mori p50(Daiz~)"
BASE COUNT      84 a      146 c      192 g      110 t
ORIGIN

Query Match      33.08; Score 25.4; DB 106; Length 532;
Best Local Similarity 61.2%; Pred. No. 50;
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 9 aatggtcaagcgttcgtctcactcgggaccgctaaagcggcccttaacaa 68
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 14 AAGCGTCCCAACGAGCGCGCTGTACTCGCTTCCGGTAGACCGCGCGGAATCCAA 73
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 69 acgttag 75
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 74 ACGTTAG 80

RESULT 11
AZ511751
LOCUS
DEFINITION lM0356K22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0356K22 R, DNA sequence.
ACCESSION AZ511751
VERSION AZ511751.1 GI:10693067
KEYWORDS GSS.

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 669)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0356 row: K column: 22
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 669.
FEATURES
    source
        1..669
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="UUGC1M0356K22"
            /clone_lib="Mouse 10kb plasmid UUGC1M library"
            /sex="Male"
            /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
            /notes="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (gil4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      167 a      158 c      113 g      231 t
ORIGIN

Query Match      32.7%; Score 25.2; DB 245; Length 669;
Best Local Similarity 60.0%; Pred. No. 61;
Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 8 aaatggttcaagcgttcgtctcactcgggaccgctaaagcggcccttaacaa 67
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 43 AACTGGTGGTTACTCTGTCTTCCCAACAGCTACAGTCCACCACACCC 102
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 68 aacgttag 77
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 AACGTTTAG 112

RESULT 12
BE187845
LOCUS BE187845 750 bp mRNA EST 25-SEP-2000
```

DEFINITION	CFCl09-R In vitro expressed cDNAs Cladosporium fulvum cDNA similar to 30 kd heat shock protein, mRNA sequence.	JOURNAL MEDLINE REFERENCE	prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
ACCESSION	BE187845	3 (sites)	20499374
VERSION	BE187845.1	REFERENCE	
KEYWORDS	EST	AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Kikuchi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Okazaki, Y., Muramatsu, M., Inoue, Y., and Hayashizaki, Y.
REFERENCE	1 (bases 1 to 750)	TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
AUTHORS	Clark, A.J., Rasmussen, S.W. and Oliver, R.P.	4 (sites)	20530913
TITLE	In vitro expressed genes of Cladosporium fulvum	JOURNAL MEDLINE REFERENCE	The RIKEN Genome Exploration Research Group Phase II Team and FANTOM Consortium.
COMMENT	Contact: R.P.Oliver Murdoch University SABC, Perth 6150, Western Australia Tel: +61-8-9360-7404 Fax: +61-8-9360-6303 Email: roliver@central.murdoch.edu.au High quality sequence stop: 750 POLYA-No.	JOURNAL	Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
FEATURES	Location/Qualifiers	5 (bases 1 to 897)	
Source	1..750		
	/organism="Cladosporium fulvum"		
	/strain="Race 4"		
	/db_xref="taxon:5499"		
	/clone_lib="In vitro expressed cDNAs"		
	/tissue_type="Mycelium"		
	/note="Vector: Lambda bluescript; A mixture of mycelial cultures grown in liquid B5 for 48 hours and transferred for 24 hours to media lacking carbon, nitrogen or supplemented with hydrogen peroxide"		
BASE COUNT	201 a 198 c 187 g 135 t 29 others		
ORIGIN			
Query Match	32.7%; Score 25.2; DB 164; Length 750;		
Best Local Similarity	61.9%; Pred. No. 62;		
Matches	39; Conservative 0; Mismatches 24; Indels 0; Gaps 0;		
QY	14 ttcaagccgttcgttcgctcactcgagccggaacgagcccttaccacaaacgtt 73		
Db	18 TGAAGCCATCCGATCGCCACAGAGACTACCTCAAGTCATCAACGCCGACCTT 77		
QY	74 aga 76		
Db	78 ACA 80		
RESULT	13		
LOCUS	AK014720 897 bp mRNA HTC 08-FEB-2001		
DEFINITION	Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched library, clone:4833418N17, full insert sequence.		
ACCESSION	AK014720		
VERSION	AK014720.1		
KEYWORDS	GI:12852739		
SOURCE	CAP trapper.		
	Mus musculus (strain:C57BL/6J) 0 day neonate head cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library		
	clone:4833418N17.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (sites)		
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Methods Enzymol. 303, 19-44 (1999)		
REFERENCE	2 (sites)		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to		
		FEATURES	Location/Qualifiers
		source	
		1..897	
		/organism="Mus musculus"	
		/strain="C57BL/6J"	
		/db_xref="taxon:10090"	
		/db_xref="MGD:MGI:1906948"	
		/db_xref="MGI:1921157"	
		/clone="4833418N17"	
		/tissue_type="head"	
		/clone_lib="RIKEN full-length enriched mouse cDNA library"	
		/dev_stage="0 day neonate"	
		209 a 221 c 200 g 267 t	
		BASE COUNT	
		ORIGIN	

```
Query Match      32.7%; Score 25.2; DB 192; Length 897;
Best Local Similarity 60.0%; Pred. No. 63;
Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy 8 aaatggtcaagcgttcgctcactcgcggaccggtcaaaagccggcccttaacca 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 AACTGGTTGTTACCTCTGTTGCTCTTCCCAACAGCTACAGTCCACACACACCC 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 68 aacgttagag 77
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 AACGTTTGTAG 485

RESULT 14
AI069008/c
LOCUS
DEFINITION      mgae0005ab05f Magnaporthe grisea Appressorium stage cDNA Library
ACCESSION      AI069008
VERSION        AI069008.1 GI:3391983
KEYWORDS        EST.
SOURCE         Magnaporthe grisea
ORGANISM       Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE      Choi,W., Fang,E., Sasinowski,H., Wing,R. and Dean,R.A.
AUTHORS        Expressed sequence characterization during appressorium formation
TITLE          in rice blast fungus, Magnaporthe grisea
JOURNAL        Unpublished (1998)
COMMENT        Contact: Dean,R.A.
                Clemson University Genomics Institute
                Clemson University
                100 Jordan Hall, Clemson University, Clemson, SC 29634, USA
                Tel: 864 656 5737
                Fax: 864 656 4293
                Email: rdean@clemson.edu
                Seq primer: T3 primer (AATTAACTCTCACTAAAGG)
                High quality sequence stop: 386.
FEATURES        Location/Qualifiers
                source
                1..852
                /organism="Magnaporthe grisea"
                /strain="70-15"
                /db_xref="taxon:148305"
                /clone="mgae0005ab05f"
                /clone_lib="Magnaporthe grisea Appressorium stage cDNA
                Library"
                /dev_stage="Germinated conidia on appressorium-inductive
                surface"
                /note="Vector: pBlueScriptII SK(+) Vector; Site1: EcoRI;
                Site2: XhoI; The appressorium formation-specific cDNA
                library was constructed from conidia germinated for 5-8
                hr on an inductive surface. The library has an average
                insert size of 1.5 kbp."
BASE COUNT      213 a 215 c 221 g 198 t 5 others
ORIGIN

Query Match      32.5%; Score 25; DB 15; Length 852;
Best Local Similarity 64.9%; Pred. No. 74;
Matches 37; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 5 aacaaatggttcagcgcgttcgctcactcgcggaccggtcaaaagccggccct 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 AACCACTGGGTCAACAGCTCTCTCGAACAGTCGGGAACGGCAAAAAGTGAACACT 371

RESULT 15
TA144C03Q
LOCUS
DEFINITION      T. brucei sheared genomic DNA clone 144c03, reverse sequence,
                genomic survey sequence.
ACCESSION      AL466861
```

```
AL466861.1 GI:11836216
GSS.
Trypanosoma brucei.
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 546)
Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajandream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrellesanger.ac.uk and
nhlsanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.
FEATURES        Location/Qualifiers
                source
                1..546
                /organism="Trypanosoma brucei"
                /strain="TREU927"
                /db_xref="taxon:5691"
                /clone="144c03"
BASE COUNT      131 a 162 c 112 g 141 t
ORIGIN

Query Match      32.2%; Score 24.8; DB 258; Length 546;
Best Local Similarity 60.3%; Pred. No. 83;
Matches 41; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 4 taacaaatggttcagcgcgttcgctcactcgcggaccggtcaaaagccggccctta 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 TAATAATCATTCACACTCATCCGACCACTTAGTCGGCATTTGGCTGCAGCGAGCTCTCA 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 64 accaaacg 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 TCCATGCG 98

Search completed: August 2, 2001, 18:38:56
Job time: 6346 sec
```



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:53:11 ; Search time 49.77 Seconds  
(without alignments)  
1081.043 Million cell updates/sec

Title: US-09-701-626A-78

Perfect score: 77  
Sequence: 1 cctcaacaatggtcaagc.....cccttaaccaaaccgttagag 77

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 327855 seqs, 349373361 residues

Total number of hits satisfying chosen parameters: 655710

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summa: ies

Database : Pending Patents\_NA\_New\*

- 1: /cgnl\_7/ptodata/2/pna/PCT\_NEW\_COMB.seq.\*
- 2: /cgnl\_7/ptodata/2/pna/US06\_NEW\_COMB.seq.\*
- 3: /cgnl\_7/ptodata/2/pna/US07\_NEW\_COMB.seq.\*
- 4: /cgnl\_7/ptodata/2/pna/US08\_NEW\_COMB.seq.\*
- 5: /cgnl\_7/ptodata/2/pna/US09\_NEW\_COMB.seq.\*
- 6: /cgnl\_7/ptodata/2/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26.4	34.3	474	5	US-09-758-472-1145
C 2	25.2	32.7	12595	5	US-09-764-902-2226
C 3	25	32.5	514	5	US-09-764-864-389
C 4	25	32.5	629	5	US-09-764-864-797
C 5	24.8	32.2	32193	5	US-09-760-471-110
C 6	24.8	32.2	32193	5	US-09-760-485-1268
C 7	24.8	32.2	32194	5	US-09-760-471-109
C 8	24.8	32.2	32194	5	US-09-760-485-1267
C 9	23.8	30.9	7727	5	US-09-764-891-7685
C 10	23.8	30.9	125803	5	US-09-803-736-621
C 11	23.4	30.4	1449	5	US-09-803-110-3192
C 12	23.4	30.4	184668	5	US-09-803-110-207
C 13	22.6	29.4	654	6	US-60-278-258-7789
C 14	22.6	29.4	2992	5	US-09-898-888-5715
C 15	22.6	29.4	55739	5	US-09-803-736-485
C 16	22.6	29.4	58690	5	US-09-803-736-209
C 17	22.6	29.4	81414	5	US-09-803-736-757
C 18	22.6	29.4	95214	5	US-09-803-736-382
C 19	22.6	29.4	110157	5	US-09-803-736-758
C 20	22.4	29.1	312	5	US-09-803-110-7068
C 21	22.4	29.1	876	5	US-09-803-110-3052
C 22	22.4	29.1	1149	5	US-09-758-439-253
C 23	22.4	29.1	3116	5	US-09-764-872-585
C 24	22.4	29.1	37604	5	US-09-803-110-194
C 25	22.4	29.1	542340	5	US-09-803-110-219
C 26	22.2	28.8	1447	5	US-09-760-475-928
C 27	22.2	28.8	1447	5	US-09-760-485-302

28 22.2 28.8 2192 6 US-60-278-258-14693 Sequence 14693, A  
29 22.2 28.8 3788 5 US-09-347-316-33 Sequence 33, Appl  
30 22.2 28.8 12269 5 US-09-760-475-3882 Sequence 3882, Ap  
31 22.2 28.8 12269 5 US-09-760-485-1308 Sequence 1308, Ap  
32 22.2 28.8 12548 5 US-09-760-475-3881 Sequence 3881, Ap  
33 22.2 28.8 90632 5 US-09-803-736-1160 Sequence 1160, Ap  
34 22 28.6 1510 5 US-09-764-891-7302 Sequence 7302, Ap  
35 22 28.6 1510 5 US-09-764-891-8373 Sequence 8373, Ap  
36 22 28.6 1510 5 US-09-764-891-8374 Sequence 8374, Ap  
37 22 28.6 32222 5 US-09-764-887-619 Sequence 619, App  
38 22 28.6 32222 5 US-09-804-730-15786 Sequence 15786, A  
39 21.8 28.3 345 5 US-09-758-474-612 Sequence 612, App  
40 21.8 28.3 104364 5 US-09-803-736-337 Sequence 337, App  
41 21.8 28.3 318095 5 US-09-803-110-215 Sequence 215, App  
42 21.6 28.1 142 5 US-09-804-730-8268 Sequence 8268, Ap  
43 21.6 28.1 474 6 US-60-255-619-14683 Sequence 14683, A  
44 21.6 28.1 626 5 US-09-757-031-90 Sequence 90, Appl  
45 21.6 28.1 626 5 US-09-757-031-90 Sequence 90, Appl

#### ALIGNMENTS

RESULT 1  
US-09-758-472-1145/c  
; Sequence 1145, Application US/09758472  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PH001  
; CURRENT APPLICATION NUMBER: US/09/758,472  
; PRIOR FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; NUMBER OF SEQ ID NOS: 9632  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1145  
; LENGTH: 474  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (382)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (442)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (453)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-758-472-1145

Query Match 34.3%; Score 26.4; DB 5; Length 474;  
Best Local Similarity 61.8%; Pred. No. 0.67; Gaps 0;  
Matches 42; Conservative 0; Mismatches 26; Indels 0;  
Qy 10 atggttaagcgttccttcgtcactcggacccggtacaaagccgccttaaccacaa 69  
|| ||||| ||||| | ||||| || ||||| ||||| |||||  
Db 131 ATAGTCTTACCGTTCGTGTAATTTCTCGTGAACTTTGATCCGCCCTCCGGCAA 72  
Qy 70 cgttagag 77  
| |||||  
Db 71 TATAAGAG 64

RESULT 2  
US-09-764-902-2226/c  
; Sequence 2226, Application US/09764902  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

```

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT213
; CURRENT APPLICATION NUMBER: US/09/764,902
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2318
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2226
; LENGTH: 12595
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-902-2226

```

```

Query Match      32.7%; Score 25.2; DB 5; Length 12595;
Best Local Similarity 60.0%; Pred. No. 4.3;
Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 3 ctacaaatggttcaagccgttcgctcactcgcggagccggtctaaagccggccctt 62
   ||||| |||| | |||| | ||||| |||| | ||||| |||| | ||||| |||| |
Db 9184 CAACAAAACCTCCTCTCTGTTGGTGGCTCCTGGAAGGAGCTAGAGCCACCCAGCG 9125

QY 63 aaccaaactg 72
   || || ||
Db 9124 AAAAAAGAGT 9115

```

```

RESULT 3
US-09-764-864-389/c
; Sequence 389, Application US/09764864
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 389
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-389

```

```

Query Match      32.5%; Score 25; DB 5; Length 514;
Best Local Similarity 58.9%; Pred. No. 2.4;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 3 ctacaaatggttcaagccgttcgctcactcgcggagccggtctaaagccggccctt 62
   |||| | |||| | |||| | ||||| |||| | |||| | |||| | |||| |
Db 371 CTACAGCTGTTGGAGCGGCGGCTTTCACGCTCTGGGGCCCGAGGCGGCGGCTG 312

QY 63 aaccaaactg 75
   |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 311 GACCAGTCGCGAG 299

```

```

RESULT 4
US-09-764-864-797
; Sequence 797, Application US/09764864
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 797
; LENGTH: 629

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (607)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (626)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (627)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (628)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (629)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-797

```

```

Query Match      32.5%; Score 25; DB 5; Length 629;
Best Local Similarity 58.9%; Pred. No. 2.5;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 3 ctacaaatggttcaagccgttcgctcactcgcggagccggtctaaagccggccctt 62
   |||| | |||| | |||| | ||||| |||| | |||| | |||| | |||| |
Db 101 ctaccagctggtgacggcggttcacgctctcgttggggcccgagggcgagctg 160

QY 63 aaccaaactg 75
   |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 161 gaccagtcgcgag 173

```

```

RESULT 5
US-09-760-471-110/c
; Sequence 110, Application US/09760471
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT240
; CURRENT APPLICATION NUMBER: US/09/760,471
; CURRENT FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 110
; LENGTH: 32193
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-760-471-110

```

```

Query Match      32.2%; Score 24.8; DB 5; Length 32193;
Best Local Similarity 67.3%; Pred. No. 7.7;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 24 tcgcttcgctcactcgcggagccggtctaaagccggcccttaaccaaactgtag 75
   || | || || | || || || || || || || || || || || || || || || ||
Db 6198 TCTCCTCCCTCCCGCGGACTGGGTAAAGCCAGCCCTGTAGCTGAATGCTAG 6147

```

```

RESULT 6
US-09-760-485-1268/c
; Sequence 1268, Application US/09760485
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ217
; CURRENT APPLICATION NUMBER: US/09/760,485
; CURRENT FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1477

```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1288
; LENGTH: 32193
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-760-485-1268

Query Match      32.2%; Score 24.8; DB 5; Length 32193;
Best Local Similarity 67.3%; Pred. No. 7.7;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 24 tcgcttcgctcactcgggacgcgttaagccggccccccttaacccaacgcttag 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6198 TCTCCTCCCTCCCGCGGACTGGGTAAAGCCAGCCCTGTAGTGAATGTAG 6147

RESULT 7
US-09-760-471-109/c
; Sequence 109, Application US/09760471
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT240
; CURRENT APPLICATION NUMBER: US/09/760,471
; CURRENT FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 109
; LENGTH: 32194
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-760-471-109

Query Match      32.2%; Score 24.8; DB 5; Length 32194;
Best Local Similarity 67.3%; Pred. No. 7.7;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 24 tcgcttcgctcactcgggacgcgttaagccggccccccttaacccaacgcttag 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7278 TCTCCTCCCTCCCGCGGACTGGGTAAAGCCAGCCCTGTAGTGAATGTAG 7227

RESULT 8
US-09-760-485-1267/c
; Sequence 1267, Application US/09760485
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ217
; CURRENT APPLICATION NUMBER: US/09/760,485
; CURRENT FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1477
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1267
; LENGTH: 32194
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-760-485-1267

Query Match      32.2%; Score 24.8; DB 5; Length 32194;
Best Local Similarity 67.3%; Pred. No. 7.7;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 24 tcgcttcgctcactcgggacgcgttaagccggccccccttaacccaacgcttag 75
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7278 TCTCCTCCCTCCCGCGGACTGGGTAAAGCCAGCCCTGTAGTGAATGTAG 7227
```

```

RESULT 9
US-09-764-891-7685/c
; Sequence 7685, Application US/09764891
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7685
; LENGTH: 7727
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7685

Query Match      30.9%; Score 23.8; DB 5; Length 7727;
Best Local Similarity 57.3%; Pred. No. 13;
Matches 43; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2 tctaacaaatggttcaagccggttcgcttcgctcactcgggacgcgttaagccgccccct 61
|| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3427 TCATACAAATCAGTCCTTTCAAATGCTTCTGTCACCTGGGCTGAGTCCAAAGGACACCTT 3468

QY 62 taaccaaacgcttaga 76
| ||||| |||
Db 3367 TCCCAAAACATTACA 3353

RESULT 10
US-09-803-736-621
; Sequence 621, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norris, Susan R.
; APPLICANT: Rounsley, Steven D.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Plant Polymorphic Markers and Uses Thereof
; FILE REFERENCE: 38-10(15493)D
; CURRENT APPLICATION NUMBER: US/09/803,736
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/534,859
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: identified by Attorney Docket number 04983.0206CPUS01 38
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 1582
; SEQ ID NO 621
; LENGTH: 125803
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-803-736-621

Query Match      30.9%; Score 23.8; DB 5; Length 125803;
Best Local Similarity 62.7%; Pred. No. 26;
Matches 37; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 6 acaaatgttcaagccggttcgcttcgctcactcgggacgcgttaagccgccccctaa 64
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52099 acaattgatcaggctgctctctatgctctctcccccacatgatgactcgggtccatta 52157

RESULT 11
US-09-803-110-3192
; Sequence 3192, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
```

```
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 3192
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-803-110-3192

Query Match          30.4%; Score 23.4; DB 5; Length 1449;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 9 aatggttaagccgttcgttcgtcactcgtggaacggcgttaagccggcccttaaccaa 68
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 391 aatgtctataccggctccgttccgcgcccggtcgtggtgaaagcgggccacttggccag 450

QY 69 acgtt 73
    || |
Db 451 ccgat 455

RESULT 12
US-09-803-110-207
; Sequence 207, Application US/09803110
; GENERAL INFORMATION:
; APPLICANT: Hinkie, Gregory J.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)D
; CURRENT APPLICATION NUMBER: US/09/803,110
; CURRENT FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/739,449
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 60/168,139
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 207
; LENGTH: 184668
; TYPE: DNA
; ORGANISM: Agrobacterium tumefaciens
US-09-803-110-207

Query Match          30.4%; Score 23.4; DB 5; Length 184668;
Best Local Similarity 60.0%; Pred. No. 40;
Matches 39; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 9 aatggttaagccgttcgttcgtcactcgtggaacggcgttaagccggcccttaaccaa 68
    |||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161493 aatgtctataccggctccgttccgcgcccggtcgtggtgaaagcgggccacttggccag 161552

QY 69 acgtt 73
    || |
Db 161553 ccgat 161557

RESULT 13
US-60-278-7789/c
; Sequence 7789, Application US/60278258
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
```

```
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; TITLE OF INVENTION: Polymorphisms Identified Thereby
; FILE REFERENCE: GX-0010-1 P
; CURRENT APPLICATION NUMBER: US/60/278,258
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 17730
; SOFTWARE: PERL Program
; SEQ ID NO 7789
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 347789.26
; NAME/KEY: unsure
; LOCATION: 179
; OTHER INFORMATION: a, t, c, g, or other
US-60-278-258-7789

Query Match          29.4%; Score 22.6; DB 6; Length 654;
Best Local Similarity 63.0%; Pred. No. 21;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 10 atggttcaagcgttcgttcgtcactcgtggaacggcgttaagccggccctta 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 AGGGTCCAGCTGGCTCAGCTCCTTCAGCTGCAGCTACAGCCAGCCNCCTTA 174

RESULT 14
US-09-898-888-5715/c
; Sequence 5715, Application US/09898888
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS cDNA
; FILE REFERENCE: 20411-748CON1
; CURRENT APPLICATION NUMBER: US/09/898,888
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 09/340,623
; PRIOR FILING DATE: 1999-06-28
; NUMBER OF SEQ ID NOS: 45207
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5715
; LENGTH: 2992
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-898-888-5715

Query Match          29.4%; Score 22.6; DB 5; Length 2992;
Best Local Similarity 55.8%; Pred. No. 30;
Matches 43; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 ctctaacaatggttcaagccgttcgttcgtcactcgtggaacggcgttaagccggcccc 60
    ||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1886 CTCACACAGATGTCCTCCTGTCAGACGCTTGACATGTCATACAGGGTGAGGGGGCCAC 1827

QY 61 ttaaccacaacgttagag 77
    | | | | | | | | | |
Db 1826 TGCAGCAGAGTCAGG 1810

RESULT 15
US-09-803-736-485
; Sequence 485, Application US/09803736
; GENERAL INFORMATION:
; APPLICANT: Bush, David F.
; APPLICANT: Levin, Irena M.
; APPLICANT: Norri, Susan R.
; APPLICANT: Rounsley, Steven D.
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using s<sub>w</sub> model  
Run on: August 2, 2001, 16:53:10 : Search time 2788.81 Seconds  
(without alignments)  
407.993 Million cell updates/sec

Title: US-09-701-626A-78  
Perfect score: 77  
Sequence: 1 ccttaacaaagtgttaagc.....cccttaacaaacgttagg 77

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 14155048 seqs, 7388405095 residues 28310096  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main:\*

1: /cgnl\_7/ptodata/1/pna/US06\_COMB.seq:\*

2: /cgnl\_7/ptodata/1/pna/US06\_COMB.seq:\*

3: /cgnl\_7/ptodata/1/pna/US07\_COMB.seq:\*

4: /cgnl\_7/ptodata/1/pna/US08\_COMB.seq:\*

5: /cgnl\_7/ptodata/1/pna/US081\_COMB.seq:\*

6: /cgnl\_7/ptodata/1/pna/US082\_COMB.seq:\*

7: /cgnl\_7/ptodata/1/pna/US083\_COMB.seq:\*

8: /cgnl\_7/ptodata/1/pna/US084\_COMB.seq:\*

9: /cgnl\_7/ptodata/1/pna/US085\_COMB.seq:\*

10: /cgnl\_7/ptodata/1/pna/US086\_COMB.seq:\*

11: /cgnl\_7/ptodata/1/pna/US087\_COMB.seq:\*

12: /cgnl\_7/ptodata/1/pna/US088\_COMB.seq:\*

13: /cgnl\_7/ptodata/1/pna/US089\_COMB.seq:\*

14: /cgnl\_7/ptodata/1/pna/US090\_COMB.seq:\*

15: /cgnl\_7/ptodata/1/pna/US091\_COMB.seq:\*

16: /cgnl\_7/ptodata/1/pna/US092\_COMB.seq:\*

17: /cgnl\_7/ptodata/1/pna/US093\_COMB.seq:\*

18: /cgnl\_7/ptodata/1/pna/US094\_COMB.seq:\*

19: /cgnl\_7/ptodata/1/pna/US095A\_COMB.seq:\*

20: /cgnl\_7/ptodata/1/pna/US095B\_COMB.seq:\*

21: /cgnl\_7/ptodata/1/pna/US095C\_COMB.seq:\*

22: /cgnl\_7/ptodata/1/pna/US095D\_COMB.seq:\*

23: /cgnl\_7/ptodata/1/pna/US096A\_COMB.seq:\*

24: /cgnl\_7/ptodata/1/pna/US096B\_COMB.seq:\*

25: /cgnl\_7/ptodata/1/pna/US096C\_COMB.seq:\*

26: /cgnl\_7/ptodata/1/pna/US096D\_COMB.seq:\*

27: /cgnl\_7/ptodata/1/pna/US096E\_COMB.seq:\*

28: /cgnl\_7/ptodata/1/pna/US097A\_COMB.seq:\*

29: /cgnl\_7/ptodata/1/pna/US097B\_COMB.seq:\*

30: /cgnl\_7/ptodata/1/pna/US097C\_COMB.seq:\*

31: /cgnl\_7/ptodata/1/pna/US098\_COMB.seq:\*

32: /cgnl\_7/ptodata/1/pna/US099\_COMB.seq:\*

33: /cgnl\_7/ptodata/1/pna/US6001\_COMB.seq:\*

34: /cgnl\_7/ptodata/1/pna/US6002\_COMB.seq:\*

35: /cgnl\_7/ptodata/1/pna/US6003\_COMB.seq:\*

36: /cgnl\_7/ptodata/1/pna/US6004\_COMB.seq:\*

37: /cgnl\_7/ptodata/1/pna/US6005\_COMB.seq:\*

38: /cgnl\_7/ptodata/1/pna/US6006\_COMB.seq:\*

39: /cgnl\_7/ptodata/1/pna/US6007\_COMB.seq:\*

40: /cgnl\_7/ptodata/1/pna/US6008\_COMB.seq:\*

41: /cgnl\_7/ptodata/1/pna/US6009\_COMB.seq:\*

42: /cgnl\_7/ptodata/1/pna/US6010\_COMB.seq:\*

43: /cgnl\_7/ptodata/1/pna/US6011\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	77	100.0	77	1	PCT-US99-13295-78
2	77	100.0	77	28	US-09-701-626A-78
3	74	96.1	77	1	PCT-US99-13295-77
4	74	96.1	77	28	US-09-701-626A-77
5	74	96.1	14143	1	PCT-US99-13295-1
6	74	96.1	14143	1	PCT-US99-13295-2
7	74	96.1	14143	28	US-09-701-626A-1
8	74	96.1	42143	1	PCT-US99-13295-4
9	71.8	93.2	77	1	PCT-US99-13295-19
10	71.8	93.2	77	28	US-09-701-626A-19
11	71.4	92.7	77	1	PCT-US99-13295-63
12	71.4	92.7	77	1	PCT-US99-13295-119
13	71.4	92.7	77	28	US-09-701-626A-63
14	71.4	92.7	77	28	US-09-701-626A-119
15	71.2	92.5	77	1	PCT-US99-13295-65
16	71.2	92.5	77	28	US-09-701-626A-65
17	69.8	90.6	77	1	PCT-US99-13295-11
18	69.8	90.6	77	28	US-09-701-626A-11
19	69.8	90.6	77	28	US-09-701-626A-24
20	69.8	90.6	77	28	US-09-701-626A-55
21	69.2	89.9	77	1	PCT-US99-13295-22
22	69.2	89.9	77	28	US-09-701-626A-22
23	68.6	89.1	77	1	PCT-US99-13295-41
24	68.6	89.1	77	28	US-09-701-626A-41
25	68.2	88.6	77	1	PCT-US99-13295-68
26	68.2	88.6	77	28	US-09-701-626A-68
27	68.2	88.6	77	28	US-09-701-626A-4
28	68.2	88.6	12956	28	US-09-701-626A-5
29	67.6	87.8	77	1	PCT-US99-13295-5
30	67.6	87.8	77	28	US-09-701-626A-5
31	67.4	87.5	77	1	PCT-US99-13295-3
32	67.4	87.5	77	28	US-09-701-626A-2
33	67.4	87.5	7300	28	US-09-701-626A-3
34	67.4	87.5	7300	28	US-09-701-626A-122
35	67.4	87.5	7300	28	US-09-701-626A-98
36	67.4	87.5	7300	28	US-09-701-626A-75
37	67.0	87.0	77	1	PCT-US99-13295-21
38	67.0	87.0	77	28	US-09-701-626A-21
39	67.0	87.0	77	28	US-09-701-626A-122
40	66.6	86.5	74	1	PCT-US99-13295-98
41	66.6	86.5	74	28	US-09-701-626A-98
42	65.4	84.9	77	1	PCT-US99-13295-75
43	65.4	84.9	77	28	US-09-701-626A-75
44	65.4	84.9	77	28	US-09-701-626A-26
45	65.4	84.9	77	1	PCT-US99-13295-26

44: /cgnl\_7/ptodata/1/pna/US6012\_COMB.seq:\*

45: /cgnl\_7/ptodata/1/pna/US6013\_COMB.seq:\*

46: /cgnl\_7/ptodata/1/pna/US6014\_COMB.seq:\*

47: /cgnl\_7/ptodata/1/pna/US6015\_COMB.seq:\*

48: /cgnl\_7/ptodata/1/pna/US6016\_COMB.seq:\*

49: /cgnl\_7/ptodata/1/pna/US6017\_COMB.seq:\*

50: /cgnl\_7/ptodata/1/pna/US6018\_COMB.seq:\*

51: /cgnl\_7/ptodata/1/pna/US6019\_COMB.seq:\*

52: /cgnl\_7/ptodata/1/pna/US6020\_COMB.seq:\*

53: /cgnl\_7/ptodata/1/pna/US6021\_COMB.seq:\*

54: /cgnl\_7/ptodata/1/pna/US6022\_COMB.seq:\*

55: /cgnl\_7/ptodata/1/pna/US6023\_COMB.seq:\*

56: /cgnl\_7/ptodata/1/pna/US6024\_COMB.seq:\*

57: /cgnl\_7/ptodata/1/pna/US6025\_COMB.seq:\*

58: /cgnl\_7/ptodata/1/pna/US6026\_COMB.seq:\*

59: /cgnl\_7/ptodata/1/pna/US6027\_COMB.seq:\*

60: /cgnl\_7/ptodata/1/pna/US6028\_COMB.seq:\*

ALIGNMENTS

```

;RESULT 1
;CT-US99-13295-78
; Sequence 78, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; PCT-US99-13295-78

Query Match 100.0%; Score 77; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.8e-19;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccttaacaaatggttcaagcgttcgctcactcgggacggcgttaaacgcccc 60
Db 1 ccttaacaaatggttcaagcgttcgctcactcgggacggcgttaaacgcccc 60

Qy 61 ttaacaaacgcttagag 77
Db 61 ttaacaaacgcttagag 77

;RESULT 2
;US-09-701-626A-78
; Sequence 78, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; US-09-701-626A-78
```

```

Query Match 100.0%; Score 77; DB 28; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.8e-19;
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccttaacaaatggttcaagcgttcgctcactcgggacggcgttaaacgcccc 60
Db 1 ccttaacaaatggttcaagcgttcgctcactcgggacggcgttaaacgcccc 60

Qy 61 ttaacaaacgcttagag 77
Db 61 ttaacaaacgcttagag 77

;RESULT 3
;PCT-US99-13295-77
; Sequence 77, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; CURRENT FILING DATE: 1999-06-11
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 77
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; PCT-US99-13295-77

Query Match 96.1%; Score 74; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 6.7e-18;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ctaacaaatggttcaagcgttcgctcactcgggacggcgttaaacgcccc 62
Db 3 ctaacaaatggttcaagcgttcgctcactcgggacggcgttaaacgcccc 62

Qy 63 aaccaaacgcttaga 76
Db 63 aaccaaacgcttaga 76

;RESULT 4
;US-09-701-626A-77
; Sequence 77, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; US-09-701-626A-77
```



; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 77  
; LENGTH: 77  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)  
US-09-701-626A-77

Query Match 96.1%; Score 74; DB 28; Length 77;  
Best Local Similarity 100.0%; Pred. No. 6.7e-18;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ctaacaaatggttaagccgttcgctcactcggtacggtctaaagccgccccctt 62  
|||||  
Db 3 ctaacaaatggttaagccgttcgctcactcggtacggtctaaagccgccccctt 62

Qy 63 aaccaaactgttaga 76  
|||||

Db 63 aaccaaactgttaga 76  
|||||

RESULT 5

US99-13295-1  
; Sequence 1, Application PC/TUS9913295  
; GENERAL INFORMATION:  
; APPLICANT: New England Biolabs, Inc.  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: Gene Discovery Method  
; CURRENT APPLICATION NUMBER: PCT/US99/13295  
; EARLIER FILING DATE: 1999-06-11  
; EARLIER FILING DATE: 1999-06-11  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER FILING DATE: 1998-06-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 14143  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Genomic DNA of Pseudomonas Alcaligenes NEB#585  
; OTHER INFORMATION: (ATCC 55044)  
PCT-US99-13295-1

Query Match 96.1%; Score 74; DB 1; Length 14143;  
Best Local Similarity 100.0%; Pred. No. 2.1e-17;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ctaacaaatggttaagccgttcgctcactcggtacggtctaaagccgccccctt 62  
|||||  
Db 9095 ctaacaaatggttaagccgttcgctcactcggtacggtctaaagccgccccctt 9154

Qy 63 aaccaaactgttaga 76  
|||||

Db 9155 aaccaaactgttaga 9168  
|||||

RESULT 6

PCT-US99-13295-2  
; Sequence 2, Application PC/TUS9913295  
; GENERAL INFORMATION:  
; APPLICANT: New England Biolabs, Inc.  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth

; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: Gene Discovery Method  
; CURRENT APPLICATION NUMBER: PCT/US99/13295  
; CURRENT FILING DATE: 1999-06-11  
; EARLIER FILING DATE: 1999-06-11  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER FILING DATE: 1998-06-12  
; EARLIER FILING DATE: 1998-06-12  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 14143  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)  
PCT-US99-13295-2

Query Match 96.1%; Score 74; DB 1; Length 14143;  
Best Local Similarity 100.0%; Pred. No. 2.1e-17;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ctaacaaatggttaagccgttcgctcactcggtacggtctaaagccgccccctt 62  
|||||  
Db 9095 ctaacaaatggttaagccgttcgctcactcggtacggtctaaagccgccccctt 9154

Qy 63 aaccaaactgttaga 76  
|||||

Db 9155 aaccaaactgttaga 9168  
|||||

RESULT 7

US-09-701-626A-1  
; Sequence 1, Application US/09701626A  
; GENERAL INFORMATION:  
; APPLICANT: Vaisvila, Romualdas  
; APPLICANT: Morgan, Richard D.  
; APPLICANT: Raleigh, Elisabeth  
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method  
; FILE REFERENCE: NEB-165PUS  
; CURRENT APPLICATION NUMBER: US/09/701,626A  
; CURRENT FILING DATE: 2001-04-16  
; PRIOR APPLICATION NUMBER: 60/089,086  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: 60/089,101  
; PRIOR FILING DATE: 1998-06-12  
; PRIOR APPLICATION NUMBER: PCT/US99/13295  
; PRIOR FILING DATE: 1999-06-11  
; NUMBER OF SEQ ID NOS: 130  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 14143  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: genomic DNA of  
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB#585 (ATCC 55044)  
US-09-701-626A-1

Query Match 96.1%; Score 74; DB 28; Length 14143;  
Best Local Similarity 100.0%; Pred. No. 2.1e-17;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 ctaacaaatggttaagccgttcgctcactcggtacggtctaaagccgccccctt 62  
|||||  
Db 9095 ctaacaaatggttaagccgttcgctcactcggtacggtctaaagccgccccctt 9154

Qy 63 aaccaaactgttaga 76  
|||||

Db 9155 aaccaaactgttaga 9168  
|||||

```
RESULT 8
PCT-US99-13295-4
; Sequence 4, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 42143
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-4

Query Match          96.1%; Score 74; DB 1; Length 42143;
Best Local Similarity 100.0%; Pred. No. 2.6e-17;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 cttaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccctt 62
DB 9095 cttaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccctt 62

QY 63 aaccacaaacgttaga 76
DB 9155 aaccacaaacgttaga 9168

RESULT 9
PCT-US99-13295-19
; Sequence 19, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-19

Query Match          93.2%; Score 71.8; DB 1; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.6e-17;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60
DB 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60

QY 61 ttaacacaaacgttag 75
DB 61 ttaacacaaacgttag 75

RESULT 10
US-09-701-626A-19
; Sequence 19, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09701.626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-19

Query Match          93.2%; Score 71.8; DB 28; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.6e-17;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60
DB 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60

QY 61 ttaacacaaacgttag 75
DB 61 ttaacacaaacgttag 75

RESULT 11
PCT-US99-13295-63
; Sequence 63, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-63

Query Match          93.2%; Score 71.8; DB 1; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.6e-17;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60
DB 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60

QY 61 ttaacacaaacgttag 75
DB 61 ttaacacaaacgttag 75
```

```
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60
DB 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60

QY 61 ttaacacaaacgttag 75
DB 61 ttaacacaaacgttag 75

RESULT 10
US-09-701-626A-19
; Sequence 19, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09701.626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
US-09-701-626A-19

Query Match          93.2%; Score 71.8; DB 28; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.6e-17;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60
DB 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60

QY 61 ttaacacaaacgttag 75
DB 61 ttaacacaaacgttag 75

RESULT 11
PCT-US99-13295-63
; Sequence 63, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER APPLICATION NUMBER: 60/089,086
; EARLIER FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/089,101
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-63

Query Match          93.2%; Score 71.8; DB 1; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.6e-17;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60
DB 1 ctctaaacaaatggttcaaacgcttcgcttcgctcactcggtgacggcgttaaacggccccc 60

QY 61 ttaacacaaacgttag 75
DB 61 ttaacacaaacgttag 75
```

```
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; PCT-US99-13295-63

Query Match          92.7%; Score 71.4; DB 1; Length 77;
Best Local Similarity 98.6%; Pred. No. 6.6e-17;
Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcggcttcgctcactcgggaccggtacaaagccggccctt 62
    |||||||
Db 3 ctaacaaatggttcaagcggcttcgctcactcgggaccggtacaaagccggccctt 62

QY 63 aaccaaaccgttag 75
    |||||||
Db 63 aaccaaaccgttag 75

RESULT 12
PCT-US99-13295-119
; Sequence 119, Application PC/TUS9913295
; GENERAL INFORMATION:
; APPLICANT: New England Biolabs, Inc.
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: Gene Discovery Method
; CURRENT APPLICATION NUMBER: PCT/US99/13295
; EARLIER FILING DATE: 1999-06-11
; EARLIER FILING DATE: 1998-06-12
; EARLIER FILING DATE: 1998-06-12
; EARLIER FILING DATE: 1998-06-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Consensus
; OTHER INFORMATION: Sequence
; PCT-US99-13295-119

Query Match          92.7%; Score 71.4; DB 1; Length 77;
Best Local Similarity 98.6%; Pred. No. 6.6e-17;
Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcggcttcgctcactcgggaccggtacaaagccggccctt 62
    |||||||
Db 3 ctaacaaatggttcaagcggcttcgctcactcgggaccggtacaaagccggccctt 62

QY 63 aaccaaaccgttag 75
    |||||||
Db 63 aaccaaaccgttag 75

RESULT 13
US-09-701-626A-63
; Sequence 63, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
```

```
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
; OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
; US-09-701-626A-63

Query Match          92.7%; Score 71.4; DB 28; Length 77;
Best Local Similarity 98.6%; Pred. No. 6.6e-17;
Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcggcttcgctcactcgggaccggtacaaagccggccctt 62
    |||||||
Db 3 ctaacaaatggttcaagcggcttcgctcactcgggaccggtacaaagccggccctt 62

QY 63 aaccaaaccgttag 75
    |||||||
Db 63 aaccaaaccgttag 75

RESULT 14
US-09-701-626A-119
; Sequence 119, Application US/09701626A
; GENERAL INFORMATION:
; APPLICANT: Vaisvila, Romualdas
; APPLICANT: Morgan, Richard D.
; APPLICANT: Raleigh, Elisabeth
; TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
; FILE REFERENCE: NEB-165PUS
; CURRENT APPLICATION NUMBER: US/09/701,626A
; CURRENT FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/089,086
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089,101
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: PCT/US99/13295
; PRIOR FILING DATE: 1999-06-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 119
; LENGTH: 77
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Consensus
; OTHER INFORMATION: Sequence
; US-09-701-626A-119

Query Match          92.7%; Score 71.4; DB 28; Length 77;
Best Local Similarity 98.6%; Pred. No. 6.6e-17;
Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcggcttcgctcactcgggaccggtacaaagccggccctt 62
    |||||||
Db 3 ctaacaaatggttcaagcggcttcgctcactcgggaccggtacaaagccggccctt 62

QY 63 aaccaaaccgttag 75
    |||||||
Db 63 aaccaaaccgttag 75
```

```

RESULT 15
PCT-US99-13295-65
: Sequence 65, Application PC/TUS9913295
: GENERAL INFORMATION:
: APPLICANT: New England Biolabs, Inc.
: APPLICANT: Valsvila, Romualdas
: APPLICANT: Morgan, Richard D.
: APPLICANT: Raleigh, Elisabeth
: TITLE OF INVENTION: Restriction Enzyme Gene Discovery Method
: FILE REFERENCE: Gene Discovery Method
: CURRENT APPLICATION NUMBER: PCT/US99/13295
: EARLIER APPLICATION NUMBER: 60/089,086
: EARLIER FILING DATE: 1998-06-12
: EARLIER APPLICATION NUMBER: 60/089,101
: EARLIER FILING DATE: 1998-06-12
: NUMBER OF SEQ ID NOS: 130
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 65
: LENGTH: 77
: TYPE: DNA
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Description of Unknown Organism: Genomic DNA of
: OTHER INFORMATION: Pseudomonas Alcaligenes NEB #585 (ATCC 55044)
PCT-US99-13295-65

```

```

Query Match          92.5%; Score 71.2; DB 1; Length 77;
Best Local Similarity 96.1%; Pred. No. 7.8e-17;
Matches 73; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tctaacaaatggttcaagccgttcgttcgtcactcgggaccggttaaagccgcccct 61
Db 2 tctaacatggttcaagccgttcgttcgtcactcgggaccggttaaagccgcccct 61

QY 62 taaccaaacggttagag 77
Db 62 taaccaaacggttagag 77

```

Search completed: August 2, 2001, 17:44:11  
Job time: 3062 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:53:10 ; Search time 1350.43 Seconds  
(without alignments)  
881.954 Million cell updates/sec

Title: US-09-701-626A-78  
Perfect score: 77  
Sequence: 1 ccttaacaatggttcaagc.....cccttaaccaaacgttagag 77

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 1344157 seqs, 7733874588 residues  
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ba3:\*  
4: gb\_in1:\*  
5: gb\_in2:\*  
6: gb\_in3:\*  
7: gb\_om:\*  
8: gb\_ov:\*  
9: gb\_pat1:\*  
10: gb\_pat2:\*  
11: gb\_ph:\*  
12: gb\_pl1:\*  
13: gb\_pl2:\*  
14: gb\_pl3:\*  
15: gb\_pl4:\*  
16: em\_ba1:\*  
17: em\_ba2:\*  
18: em\_fun:\*  
19: em\_higo\_hum:\*  
20: em\_higo\_inv:\*  
21: em\_higo\_rod:\*  
22: em\_hig\_hum1:\*  
23: em\_hig\_hum2:\*  
24: em\_hig\_hum3:\*  
25: em\_hig\_hum4:\*  
26: em\_hig\_hum5:\*  
27: em\_hig\_hum6:\*  
28: em\_hig\_hum7:\*  
29: em\_hig\_hum8:\*  
30: em\_hig\_inv1:\*  
31: em\_hig\_inv2:\*  
32: em\_hig\_other:\*  
33: em\_hig\_rod:\*  
34: em\_hum1:\*  
35: em\_hum2:\*  
36: em\_hum3:\*  
37: em\_hum4:\*  
38: em\_hum5:\*  
39: em\_hum6:\*  
40: em\_hum7:\*  
41: em\_in:\*  
42: em\_om:\*  
43: em\_or:\*

44: em\_ov:\*  
45: em\_pat:\*  
46: em\_ph:\*  
47: em\_pl:\*  
48: em\_ro:\*  
49: em\_sts:\*  
50: em\_sy:\*  
51: em\_un:\*  
52: em\_v1:\*  
53: gb\_stsl:\*  
54: gb\_sts2:\*  
55: gb\_sts3:\*  
56: gb\_sy:\*  
57: gb\_un:\*  
58: gb\_vil:\*  
59: gb\_vil2:\*  
60: gb\_htg1:\*  
61: gb\_htg2:\*  
62: gb\_htg3:\*  
63: gb\_htg4:\*  
64: gb\_htg5:\*  
65: gb\_htg6:\*  
66: gb\_htg7:\*  
67: gb\_htg8:\*  
68: gb\_htg9:\*  
69: gb\_htg10:\*  
70: gb\_htg11:\*  
71: gb\_htg12:\*  
72: gb\_htg13:\*  
73: gb\_htg14:\*  
74: gb\_htg15:\*  
75: gb\_htg16:\*  
76: gb\_htg17:\*  
77: gb\_htg18:\*  
78: gb\_htg19:\*  
79: gb\_htg20:\*  
80: gb\_htg21:\*  
81: gb\_htg22:\*  
82: gb\_htg23:\*  
83: gb\_htg24:\*  
84: gb\_htg25:\*  
85: gb\_pr1:\*  
86: gb\_pr2:\*  
87: gb\_pr3:\*  
88: gb\_pr4:\*  
89: gb\_pr5:\*  
90: gb\_pr6:\*  
91: gb\_pr7:\*  
92: gb\_pr8:\*  
93: gb\_pr9:\*  
94: gb\_rol:\*  
95: gb\_rod:\*  
96: gb\_in4:\*  
97: gb\_pr10:\*  
98: em\_ba3:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	40.4	52.5	9581	3	PAE223604	Pseudomon
2	30.2	39.2	174732	62	AC012160	AC012160 Drosophil
3	26.8	34.8	179936	69	AC025491	AC025491 Homo sapi
4	26.6	34.5	11014	3	MTPFZ1A	X68367 M.thermofo
5	26.4	34.3	12249	1	AE000501	AE000501 Escherich
6	26.4	34.3	28446	66	AC020814	AC020814 Mus muscu
7	26.4	34.3	160312	62	AC011073	AC011073 Homo sapi
8	26.4	34.3	181215	86	AC006461	AC006461 Homo sapi

```

9      26.4 34.3 338534 2  ECUW93
c 10    26 33.8 181931 79  AL353773
c 11    26 33.8 195032 86  AC006239 Homo sapi
c 12    25.8 33.5 195000 43  AP000396 Beta vulg
c 13    25.2 32.7 2051 2  AY007784 Acinetoba
c 14    25.2 32.7 10521 65  AC020168 Drosophill
c 15    25.2 32.7 37027 89  AL135844 Human DNA
c 16    25.2 32.7 142229 30  AC007753 Drosophill
c 17    25.2 32.7 184272 4  AC009211 Drosophill
c 18    25.2 32.7 196928 82  AP000847 Homo sapi
c 19    25.2 32.7 225668 5  AE003699 Drosophill
c 20    25 32.5 199380 72  AC048379 Homo sapi
c 21    25 32.5 210000 60  AC004555 Homo sapi
c 22    24.8 32.2 828 2  AF217619 Mesorhizo
c 23    24.8 32.2 6502 85  AB038980 Homo sapi
c 24    24.8 32.2 10945 1  AE006262 Lactococc
c 25    24.8 32.2 14466 1  AE005662 Escherich
c 26    24.8 32.2 151840 69  AC026226 Homo sapi
c 27    24.8 32.2 168764 2  AP002569 Escherich
c 28    24.8 32.2 176919 67  AC023169 Homo sapi
c 29    24.8 32.2 182178 60  AC007256 Homo sapi
c 30    24.8 32.2 182309 75  AC074016 Homo sapi
c 31    24.4 31.7 87258 88  AF091512 Homo sapi
c 32    24.4 31.7 110000 84  LMFLCHR26_3
c 33    24.4 31.7 141832 91  AP000432 Homo sapi
c 34    24.4 31.7 151097 76  AC083846 Homo sapi
c 35    24.4 31.7 152044 73  AC088255 Homo sapi
c 36    24.4 31.7 168700 61  AC009115 Homo sapi
c 37    24.4 31.7 196055 69  AC025513 Homo sapi
c 38    24.4 31.7 239566 93  HSAJ03147 Homo sapi
c 39    24.4 31.7 340000 91  AP001670 Homo sapi
c 40    24.2 31.4 473 1  ABA251519 Acinetoba
c 41    24.2 31.4 3407 3  RSP243600 Rhodobact
c 42    24.2 31.4 118795 60  AC008618 Homo sapi
c 43    24.2 31.4 167086 86  AC008565 Homo sapi
c 44    24.2 31.4 171909 70  AC026470 Homo sapi
c 45    24.2 31.4 173396 80  AL390236 Homo sapi

```

## ALIGNMENTS

```

RESULT 1
PAE223604 PAE223604 9581 bp DNA BCT 19-JUL-2000
LOCUS Pseudomonas aeruginosa integron In 101 DNA.
DEFINITION AJ223604
ACCESSION AJ223604.1 GI:4210822
VERSION AJ223604.1
KEYWORDS Pseudomonas aeruginosa.
SOURCE Pseudomonas aeruginosa.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 9581)
AUTHORS Laraki,N., Galleni,M., Thamm,I., Riccio,M.L., Amicosante,G.,
Freire,J.M., and Rossolini,G.M.
TITLE Structure of In31, a blaIMP-containing Pseudomonas aeruginosa
integron phylogenically related to In5, which carries an unusual
array of gene cassettes
JOURNAL Antimicrob. Agents Chemother. 43 (4), 890-901 (1999)
MEDLINE 99216901
REFERENCE 2 (bases 1 to 9581)
AUTHORS Galleni,M.G.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-1998) Galleni M.G., Centre for Protein
Engineering (CIP), University of Liege, B6 Sart Tilman, Liege,
Liege B4000, Belgium
FEATURES Location/Qualifiers
source 1..9581
/organism="Pseudomonas aeruginosa"
repeat_region 117..140
/db_xref="taxon:287"
/feature="IRI"

```

```

/rpt_type-INVERTED
117..1467
/note="5' conserved element"
complement(319..1332)
/gene="intI1"
complement(319..1332)
/gene="intI1"
/codon_start-1
/transl_table-1
/protein_id="CA11470.1"
/db_xref="GI:4210823"
/db_xref="SPTREMBL:Q9R819"
/translating="MKTATAPLPLRSVKVLDOLRRIRVLYSLRTEQAVVHVRAF
IRHGVRHPATLGSSEVEAFSLWLANERKYSVSTHROALAAALLFYKVLCTDLPWLQ
EIGRPRSRRLPVVLPDVEVRIILGFLGEHRLFAQLLYGTGMRISGLQRLVKDLQ
DHGTIIVREGKSKDRALMLPESLAPTLRQLSRARAWLKDQAEHSGVALPALER
KYPRAGHSWFWFWFAQHTSTDPSCGVVRHHMYDQTFQFAEKRAVEQAGITKPATP
HILRHSPATALLRSYDVRTVQDLLGHSVSTTMYTHVLKVGAGVRSPLDALPPLT
SER"
1468..2346
/gene="bla-imp"
1468..2346
/gene="bla-imp"
1485..2225
/gene="bla-imp"
/codon_start-1
/transl_table-1
/protein_id="CA11471.1"
/db_xref="GI:4210824"
/translating="MSKLSVFFIFLFCSIATAAESLPDLKTEKIDEGVYVHTSFEVNV
GWGVPHGLVNLVNAEAYLIDTPFTAKDTKLVTFWVERGYKIKGSISSHFHSDSTG
GIWLNRSIPTVASELTNELLKKDKGKQVATNSFGVNYLWVKNKIEVYPCGHTPD
NVVWLPERKILFGCCFIKPIYGLNGLGDANIEAWPKSAKLLKSKYKAKLKVPPSHSEV
GDASLLKLTLEQAVKGLNESKPKSPSN"
2262..2346
/gene="bla-imp"
/note="59 bp element"
2347..2984
/gene="aacA4"
2347..2984
/gene="aacA4"
2371..2925
/gene="aacA4"
/codon_start-1
/transl_table-1
/protein_id="CA11472.1"
/db_xref="GI:4210825"
/db_xref="SPTREMBL:Q51410"
/translating="MTNSTDSVTLLRMTLMEHDLAMLYEWLNRSIHVEMWGGEARPTLA
DVORQYLPVLAQESVTPYIAMLNGEPYIGVYALGSGDGMWEEETDPCVRGIDQL
LANASOLGKGLTKLVRALVELLFNDPEVTKIQTPSPSNLRAIRCEYKAGFERQGTV
TTPDPAVYVQTRQAFERTSDA"
2920..2984
/gene="aacA4"
/note="59 bp element"
2985..3715
/gene="catB6"
2985..3715
/gene="catB6"
3018..3650
/gene="catB6"
/codon_start-1
/transl_table-1
/protein_id="CA11473.1"
/db_xref="GI:4210826"
/db_xref="SPTREMBL:Q9R818"
/translating="MENYFDSPPFKGLLSQVNTNRNKKVGRYSYSGYHGHSDDDCA
RYLLPDRDDVDKLIIGSFCISGASFIAGNQGRHWDVTSFPPFYQEPAPFSST
DAFOKAGDSITVNDVWIGISEAMIPGKIGDVGIVRSVLTVDVEPTIIGGNPAQ
IKRKFSDSEISLLMEMEMWNNWPLDKIKTAPLLCCSDIFGLHRRHWRGI"
3645..3715
/gene="catB6"
/note="59 bp element"

```



Rubin,G.M.  
Direct Submission  
Submitted (21-OCT-1999), Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Nov 24, 1999 this sequence version replaced gi:6226995.  
For further information about this sequence, please visit our location  
and relationship to other sequences, including its location  
archive web site (<http://www.fruitfly.org/sequence/>) or send email  
to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 93 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 572: contig of 572 bp in length  
573 652: gap of unknown length  
574 1157: contig of 505 bp in length  
575 1158 1237: gap of unknown length  
576 1238 1708: contig of 471 bp in length  
577 1709 1788: gap of unknown length  
578 1789 2243: contig of 461 bp in length  
579 2250 2329: gap of unknown length  
580 2330 3509: contig of 1180 bp in length  
581 3510 3589: gap of unknown length  
582 3590 4696: contig of 1107 bp in length  
583 4697 4776: gap of unknown length  
584 4777 5387: contig of 611 bp in length  
585 5388 5467: gap of unknown length  
586 5468 5971: contig of 504 bp in length  
587 5972 6051: gap of unknown length  
588 6052 6679: contig of 628 bp in length  
589 6680 6759: gap of unknown length  
590 6760 7223: contig of 464 bp in length  
591 7224 7303: gap of unknown length  
592 7304 7890: contig of 587 bp in length  
593 7891 7970: gap of unknown length  
594 7971 8553: contig of 583 bp in length  
595 8554 8633: gap of unknown length  
596 8634 9329: contig of 696 bp in length  
597 9330 9409: gap of unknown length  
598 9410 10185: contig of 776 bp in length  
599 10186 10265: gap of unknown length  
600 10266 11634: contig of 1369 bp in length  
601 11635 11714: gap of unknown length  
602 11715 12478: contig of 764 bp in length  
603 12479 12558: gap of unknown length  
604 12559 14183: contig of 1625 bp in length  
605 14184 14263: gap of unknown length  
606 14264 15048: contig of 785 bp in length  
607 15049 15128: gap of unknown length  
608 15129 16098: contig of 968 bp in length  
609 16099 16176: gap of unknown length  
610 16177 17599: contig of 1423 bp in length  
611 17600 17660 17679: gap of unknown length  
612 17680 19313: contig of 1634 bp in length  
613 19314 19393: gap of unknown length  
614 19394 20834: contig of 1441 bp in length  
615 20835 20914: gap of unknown length  
616 20915 22432: contig of 1518 bp in length  
617 22433 22512: gap of unknown length  
618 22513 24093: contig of 1581 bp in length  
619 24094 24173: gap of unknown length  
620 24174 25950: contig of 1777 bp in length  
621 25951 26030: gap of unknown length  
622 26031 27724 27803: contig of 1693 bp in length  
623 27724 29349: gap of unknown length  
624 29350 30829: gap of unknown length  
625 29430 30829: contig of 1400 bp in length

30830 30909: gap of unknown length  
30910 32162: contig of 1253 bp in length  
32163 32242: gap of unknown length  
32243 33213: contig of 971 bp in length  
32244 33293: gap of unknown length  
32294 34919: contig of 1626 bp in length  
34920 34999: gap of unknown length  
35000 37882: contig of 2783 bp in length  
37883 37862: gap of unknown length  
37863 39272: contig of 1410 bp in length  
39273 39352: gap of unknown length  
39353 40852: contig of 1500 bp in length  
40853 40932: gap of unknown length  
40933 43628: contig of 2696 bp in length  
43629 43708: gap of unknown length  
43709 45751: contig of 2043 bp in length  
45752 45831: gap of unknown length  
45832 48055: contig of 2224 bp in length  
48056 48135: gap of unknown length  
48136 49916: contig of 1781 bp in length  
49917 49996: gap of unknown length  
49997 53377: contig of 3381 bp in length  
53378 53457: gap of unknown length  
53458 57017: contig of 3560 bp in length  
57018 57097: gap of unknown length  
57098 59417: contig of 2320 bp in length  
59418 59497: gap of unknown length  
59498 63559: contig of 4062 bp in length  
63560 63639: gap of unknown length  
63640 68049: contig of 4410 bp in length  
68050 68129: gap of unknown length  
68130 73531: contig of 5402 bp in length  
73532 73611: gap of unknown length  
73612 79226: contig of 5615 bp in length  
79227 79306: gap of unknown length  
79307 84116: contig of 4810 bp in length  
84117 91839: contig of 7643 bp in length  
91840 91919: gap of unknown length  
91920 98216: contig of 6297 bp in length  
98217 98296: gap of unknown length  
98297 106059: contig of 7763 bp in length  
106060 106139: gap of unknown length  
106140 112697: contig of 6558 bp in length  
112698 112777: gap of unknown length  
112778 120287: contig of 7510 bp in length  
120288 120367: gap of unknown length  
120368 128442: contig of 8075 bp in length  
128443 128522: gap of unknown length  
128523 139563 139642: gap of unknown length  
139643 150197: contig of 11040 bp in length  
150198 150277: gap of unknown length  
150278 150656: contig of 379 bp in length  
150657 150736: gap of unknown length  
150737 151120: contig of 384 bp in length  
151121 151200: gap of unknown length  
151201 151836: contig of 636 bp in length  
151837 151916: gap of unknown length  
151917 152572: contig of 656 bp in length  
152573 152652: gap of unknown length  
152653 153265: contig of 613 bp in length  
153266 153345: gap of unknown length  
153346 153442: contig of 497 bp in length  
153443 153922: gap of unknown length  
153923 155167: contig of 1245 bp in length  
155168 155247: gap of unknown length  
155248 155865: contig of 618 bp in length  
155866 155945: gap of unknown length  
155946 156581: contig of 636 bp in length  
156582 156661: gap of unknown length  
156662 157131: contig of 470 bp in length  
157132 157211: gap of unknown length



Mon Aug 6 10:48:46 2001

```

* 157212 157702: contig of 491 bp in length
* 157703 157782: gap of unknown length
* 157783 158312: contig of 530 bp in length
* 158313 158392: gap of unknown length
* 158393 159013: contig of 621 bp in length
* 159014 159093: gap of unknown length
* 159094 159952: contig of 859 bp in length
* 159953 160032: gap of unknown length
* 160033 160835: contig of 803 bp in length
* 160836 160915: gap of unknown length
* 160916 161617: contig of 702 bp in length
* 161618 161697: gap of unknown length
* 161698 162146: contig of 449 bp in length
* 162147 162226: gap of unknown length
* 162227 162842: contig of 616 bp in length
* 162843 162922: gap of unknown length
* 162923 163348: contig of 426 bp in length
* 163349 163428: gap of unknown length
* 163429 163916: contig of 488 bp in length
* 163917 163996: gap of unknown length
* 163997 164860: contig of 864 bp in length
* 164861 164940: gap of unknown length

Query Match      39.2% Score 30.2; DB 62; Length 174732;
Best Local Similarity 69.5%; Pred. No. 1 4;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 13 gttcaagccgttcctgcctacgcgggacgcggtaaacgcggcccttaacccaagc 71
||||| ||||| || |||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
Db 18837 GTTCAGCCGTTGCTCGCTCTCCGCGGGGACCAACACCGCCACCTTAACCAAGG 18779

RESULT 3
AC025491
LOCUS      AC025491 179936 bp DNA HTG 28-MAR-2000
DEFINITION Homo sapiens chromosome 6 clone RP11-196B15 map 6, WORKING DRAFT
SEQUENCE 29 unordered pieces.
ACCESSION AC025491
VERSION   AC025491.2 GI:7331526
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179936)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
           Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
           Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
           Campoliano,A., Castie,A., Choepel,Y., Colangelo,M., Collins,S.,
           Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
           Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
           Galagan,J., Gardyna,S., Glnde,S., Goette,M., Graham,L.,
           Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
           Klein,J., Lakoque,K., Lemazares,R., Landers,T., Lehocsky,J.,
           Levine,R., Liu,G., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
           McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
           Melidrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
           Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
           Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
           Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
           Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
           Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
           Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
           Young,G., Zainoun,J., Zimmer,A. and Zody,M.
           Direct Submission
           Submitted (09-MAR-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA

```

## COMMENT

```

On Mar 28, 2000 this sequence version replaced gi:7212075.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7649
Center Clone name: 196_B_15
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 163781 bases at least Q40
Consensus quality: 171238 bases at least Q30
Consensus quality: 174757 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 177136; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1579: contig of 1579 bp in length
* 1580 1679: gap of 100 bp
* 1680 3543: contig of 1864 bp in length
* 3544 3643: gap of 100 bp
* 3644 5682: contig of 2039 bp in length
* 5683 5782: gap of 100 bp
* 5783 7667: contig of 1885 bp in length
* 7668 7767: gap of 100 bp
* 7768 10500: contig of 2733 bp in length
* 10501 10600: gap of 100 bp
* 10601 12890: contig of 2290 bp in length
* 12891 12990: gap of 100 bp
* 12991 14880: contig of 1890 bp in length
* 14881 14980: gap of 100 bp
* 14981 17380: contig of 2400 bp in length
* 17381 17480: gap of 100 bp
* 17481 20807: contig of 3327 bp in length
* 20808 20907: gap of 100 bp
* 20908 23928: contig of 3021 bp in length
* 23929 24028: gap of 100 bp
* 24029 27543: contig of 3515 bp in length
* 27544 27643: gap of 100 bp
* 27644 31065: contig of 3422 bp in length
* 31066 31165: gap of 100 bp
* 31166 35411: contig of 4246 bp in length
* 35412 35511: gap of 100 bp
* 35512 39412: contig of 3901 bp in length
* 39413 39512: gap of 100 bp
* 39513 43164: contig of 3652 bp in length
* 43165 43264: gap of 100 bp
* 43265 46755: contig of 3491 bp in length
* 46756 46855: gap of 100 bp
* 46856 53117: contig of 6262 bp in length
* 53118 53217: gap of 100 bp
* 53218 57708: contig of 4491 bp in length
* 57709 57808: gap of 100 bp
* 57809 62533: contig of 4725 bp in length
* 62534 62633: gap of 100 bp
* 62634 68232: contig of 5599 bp in length
* 68233 68332: gap of 100 bp
* 68333 75673: contig of 7341 bp in length

```





LOCUS	AE000501	12249 bp	DNA	BCT	01-DEC-2000
DEFINITION	Escherichia coli K12 MG1655 section 391 of 400 of the complete genome.				
ACCESSION	AE000501	U00096			
VERSION	AE000501.1	GI:2367372			
KEYWORDS					
SOURCE	Escherichia coli K12.				
ORGANISM	Escherichia coli K12. Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				
REFERENCE	1 (bases 1 to 12249) Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.P., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y. The complete genome sequence of Escherichia coli K-12 Science 277 (5331), 1453-1474 (1997)				
JOURNAL	97426617				
MEDLINE	9278503				
PUBMED					
REFERENCE	2 (bases 1 to 12249) Blattner,F.R. Direct Submission Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459				
JOURNAL					
REFERENCE	3 (bases 1 to 12249) Blattner,F.R. Direct Submission Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459				
JOURNAL					
REFERENCE	4 (bases 1 to 12249) Plunkett,G. III. Direct Submission Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA On Sep 9, 1997 this sequence version replaced gi:1790756. This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible ( <a href="http://cgsc.biology.yale.edu">http://cgsc.biology.yale.edu</a> ). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site ( <a href="http://www.genetics.wisc.edu">http://www.genetics.wisc.edu</a> ). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.				
FEATURES	Location/Qualifiers				
source	1..12249 /organism="Escherichia coli K12" /strain="K12" /sub_strain="MG1655" /db_xref="taxon:83333" complement(96..902)				
gene					

CDS	/gene="sgcQ"	/note="b4303"
	complement(96..902)	
	/gene="sgcQ"	
	/function="putative enzyme; Not classified"	
	/note="f268a; formerly designated yjhm"	
	/codon_start=1	
	/transl_table=11	
	/product="putative nucleoside triphosphatase"	
	/protein_id="AAC77259.1"	
	/db_xref="GI:1790757"	
	/translation="MSWLKEVIGTEKAVIAMCHLRALPGDPSFDAQLGMMNVIDKAMD DMLAQGVDAVMFNEFSPLTKVTPETTAAMARIIGLMSDIRIPFGVNVLMDDP VASFDLMAATKIREFTGAYASDFGVMDTYNETIRHQHRIAGAGVKTLFNIVPE AAVYLGNRDICSIAKSTVNNHPDALCVSGLTAGTRTDSALLKRVKTVPDIVVLANT GCVLNVPEOLSIADGCVTATTFKKDGVFANFVQARVYQSFMEKVHHR"	
	complement(905..931)	
promoter	/note="factor Sigma70; predicted +1 start at 4526481"	
gene	complement(915..2228)	
	/gene="sgcC"	
	/note="b4304"	
CDS	complement(915..2228)	
	/gene="sgcC"	
	/function="putative transport; Not classified"	
	/note="f437; formerly designated yjhn"	
	/codon_start=1	
	/transl_table=11	
	/product="putative Pts system enzyme IIC component"	
	/protein_id="AAC77260.1"	
	/db_xref="GI:1790758"	
	/translation="MEDYILSLGGTVFVPIIMVIGLIFRIPMLQAIKAGVTVGIGFV GMGLIVMAIDSLSPPIKMIERFGLALHFDVGAGAGSYGATACAMILIPVIFLL NVAMLVTRLTMMNVDIYWHYAITGTVOLMTGSLYGVLGATCHAALSLLKMDWT AKRVONIVGLEISIPQGYGSSVPLFVLLDAIVEKIPFMRGRNIDAEIOKRYGMVG DPVILGVYGLIFLAGEGEGKGCASLMTVAAMVLPFMRIRLIVELGSLIPSDGARK FFQYFKGREYVIGLDTAVTLGHPTLTAVGULLLIPMLILASILDPGNVLPDLPLVA PFFTCMATVTHRGDLVRLTISGVIVMTVLLIATQFAPYFTEMLKGGFSFAGESAQI SALSVGNMFGMSISLMSLSLGIIGVVAVGIVASVVVLFRLKRKRELSE"	
	complement(923..950)	
promoter	/gene="sgcC"	
	/note="factor Sigma70; predicted +1 start at 4526499"	
promoter	complement(952..982)	
	/gene="sgcC"	
	/note="factor Sigma70; predicted +1 start at 4526528"	
promoter	complement(960..987)	
	/gene="sgcC"	
	/note="factor Sigma70; predicted +1 start at 4526536"	
promoter	complement(988..1015)	
	/gene="sgcC"	
	/note="factor Sigma70; predicted +1 start at 4526564"	
promoter	complement(1046..1075)	
	/gene="sgcC"	
	/note="factor Sigma70; predicted +1 start at 4526622"	
gene	complement(2515..3666)	
	/gene="sgcX"	
	/note="b4305"	
CDS	complement(2515..3666)	
	/gene="sgcX"	
	/function="putative enzyme; Not classified"	
	/note="f383; formerly designated yjho"	
	/codon_start=1	
	/transl_table=11	
	/product="putative lyase/synthase"	
	/protein_id="AAC77261.1"	
	/db_xref="GI:2367373"	
	/translation="MIVLRSTEIAMSFSVQETLFSLLQHNNAISGHENAVADVMLCEFR RQAEVWRDRLGNVARYGSKDPLRMIFAHMDEVGFMYRKIPEPSGLFRFERYGGP AQVTMAGSIVTLTGDKGPGMCGICKSYHFAGKGDERTQSPSVKLMIDIGAKDKDAI RMQVGTGTPVLYNPPOLLANDLVCSKALDRLGCTALLGVADAISTMELDAVYLVVA SVQEEFNIRGIVPVLRVRKPDLAGIDITPSCDTPDLHDYSEVRINOQGVGTCLNYHG RGTLAGLITPRLIRMLEQTALEHNIPVQREVAFCVITETGYIQVEQGIQPCASLSIP CRYTHSPAENVASLRDLTDCITRLLTALAGMSAAHFPVPEPDSTGTTQEAHPL"	
	complement(3819..3848)	
promoter		

```
/note="factor Sigma70; predicted +1 start at 4529395"
complement(4422..5168)
/genes="yjhq"
/feature="b4306"
complement(4422..5168)
/genes="yjhq"
/function="putative enzyme; Not classified"
/note="f248; 100 pct identical amino acid sequence and
equal length to YJHP_ECOLI SW: P39367"
/codon_start=1
/transl_table=11
/product="putative methyltransferase"
/protein_id="AAC77262.1"
/db_xref="GI:1790760"
/translation="MDPRFTTISEHRIHNPFTTEKYATLGRVLRMKPGTRILDIG
SGSEMLCTWARDHGITGTIDMSLSFTAQAKRAEELGVSERVHFTHNDAAQVYANE
KSDVAACVGTWAGTGAAGAGAEALLAOSLPGGIMLIGEPYWRQLPATPEELAOACGVSS
TDFLTLPLGLVAGFDDLDGYDVEMVLADQEGDWYEAAKWLTHRWLEANPDDDFPARE
VRAELNAPRYVYARECFGWGVFALIR"
complement(5170..5201)
/note="factor Sigma70; predicted +1 start at 4530746"
complement(5216..5246)
/note="factor Sigma70; predicted +1 start at 4530792"
complement(5224..5769)
/genes="yjhq"
/note="b4307"
complement(5224..5769)
/genes="yjhq"
/function="orf; Unknown"
/note="f181; 100 pct identical amino acid sequence and
equal length to YJHQ_ECOLI SW: P39368"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC77263.1"
/db_xref="GI:1790761"
/translation="MTVHFTFHITKSDASDIREVETRAFSGKSLADLVASLLEDES
ARPLSLLARYEGKAVGHILFTFRKGDMSPLMHILAPLAVIPEYQGMGVGRLIR
TGIEHLRLMGCOFVFLGHATYPRHGFEPGADKGYAPYPIPEEHKACMMQSLTA
QPMTLTGHCRCADPDDETALT"
complement(5376..5391)
/protein_bind
/genes="yjhq"
/note="central position to predicted promoter: -206"
/bound_moiety="Fh1A predicted site"
complement(5376..5391)
/genes="yjhq"
/note="central position to predicted promoter: -160"
/bound_moiety="Fh1A predicted site"
complement(5787..5816)
/note="factor Sigma70; predicted +1 start at 4531363"
complement(5806..5835)
/note="factor Sigma70; predicted +1 start at 4531382"
6885..6916
/note="factor Sigma70; predicted +1 start at 4532506"

Query Match 34.3%; Score 26.4; DB 1; Length 12249;
Best Local Similarity 61.8%; Pred. No. 39;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 10 atgttcaagccttcgctcactcgggaccggtacgaacgagccgccttaaccaaa 69
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 9866 ATAGTCTTACCGTCTCGGTAATTTCTGCTGAACCTTTGAATCCGCCCTCGGCAA 9925
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 70 cggttagag 77
| | | | |
Db 9926 TATAAGAG 9933

RESULT 6
AC020814/c
LOCUS
DEFINITION Mus musculus clone RG-MBAC_270F3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC020814
```

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC020814.2 GI:6980191  
HTG; HTGS\_PHASE0.  
house mouse.  
Mus musculus  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 28446)  
DOE Joint Genome Institute.  
Sequencing of Mouse  
Unpublished  
2 (bases 1 to 28446)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 16, 2000 this sequence version replaced gi:6686494.  
\* NOTE: This record contains 28 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 602: contig of 602 bp in length  
\* gap of unknown length  
\* 603 1582: contig of 980 bp in length  
\* gap of unknown length  
\* 1583 1960: contig of 378 bp in length  
\* gap of unknown length  
\* 1961 2981: contig of 1021 bp in length  
\* gap of unknown length  
\* 2982 3775: contig of 794 bp in length  
\* gap of unknown length  
\* 3776 4866: contig of 1091 bp in length  
\* gap of unknown length  
\* 4867 5740: contig of 874 bp in length  
\* gap of unknown length  
\* 5741 6748: contig of 1008 bp in length  
\* gap of unknown length  
\* 6749 6915: contig of 167 bp in length  
\* gap of unknown length  
\* 6916 8285: contig of 1370 bp in length  
\* gap of unknown length  
\* 8286 9351: contig of 1066 bp in length  
\* gap of unknown length  
\* 9352 10403: contig of 1052 bp in length  
\* gap of unknown length  
\* 10404 12170: contig of 1767 bp in length  
\* gap of unknown length  
\* 12171 13169: contig of 999 bp in length  
\* gap of unknown length  
\* 13170 14130: contig of 961 bp in length  
\* gap of unknown length  
\* 14131 14556: contig of 426 bp in length  
\* gap of unknown length  
\* 14557 15627: contig of 1071 bp in length  
\* gap of unknown length  
\* 15628 16621: contig of 994 bp in length  
\* gap of unknown length  
\* 16622 17204: contig of 583 bp in length  
\* gap of unknown length  
\* 17205 18280: contig of 1076 bp in length  
\* gap of unknown length  
\* 18281 19621: contig of 1341 bp in length  
\* gap of unknown length  
\* 19622 20781: contig of 1160 bp in length  
\* gap of unknown length  
\* 20782 22156: contig of 1375 bp in length  
\* gap of unknown length



```

DEFINITION Homo sapiens BAC clone RP11-343N14 from 2, complete sequence.
ACCESSION AC006461
VERSION AC006461.2 GI:4508128
KEYWORDS HTG.
SOURCE human.
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 181215)
  Kalicki J., Bauer C. and Smith R.
  The sequence of Homo sapiens BAC clone RP11-343N14
  Unpublished
  2 (bases 1 to 181215)
  Waterston R.H.
  Direct Submission
  Submitted (28-JAN-1999) Genome Sequencing Center, Washington
  University School of Medicine, 4444 Forest Park Parkway, St. Louis,
  MO 63108, USA
  3 (bases 1 to 181215)
  Waterston R.H.
  Direct Submission
  Submitted (24-MAR-1999) Genome Sequencing Center, Washington
  University School of Medicine, 4444 Forest Park Parkway, St. Louis,
  MO 63108, USA
  4 (bases 1 to 181215)
  Waterston R.H.
  Direct Submission
  Submitted (22-OCT-1999) Department of Genetics, Washington
  University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
  5 (bases 1 to 181215)
  Waterston R.H.
  Direct Submission
  Submitted (21-DEC-1999) Department of Genetics, Washington
  University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
  On Mar 24, 1999 this sequence version replaced gi:4199979.
  ----- Genome Center
  Center: Washington University Genome Sequencing Center
  Center code: WUGSC
  Web site: http://genome.wustl.edu/gsc
  Contact: sapiense@wustl.wustl.edu
  ----- Summary Statistics
  -----
  NOTICE: This sequence may not represent the entire insert of this
  clone. It may be shorter because we only sequence overlapping
  clone sections once, or longer because we provide a small overlap
  between neighboring data submissions.

  This sequence was finished as follows unless otherwise noted:
  all regions were double stranded, sequenced with an alternate
  chemistry, or covered by high quality data (i.e., phred quality >=
  30); an attempt was made to resolve all sequencing problems, such
  as compressions and repeats; all regions were covered by sequence
  from more than one subclone; and the assembly was confirmed by
  restriction digest.

  MAPPING INFORMATION:
  Mapping information for this clone was provided by Dr. John D.
  McPherson, Department of Genetics, Washington University, St. Louis
  MO. For additional information about the map position of this
  sequence, see http://genome.wustl.edu/gsc

  SOURCE INFORMATION:
  The RPC1-11 human BAC library was made from the blood of one male
  donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
  Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
  approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Plier de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACE3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-493L16. Actual start of
this clone is at base position 1 of RP11-343N14; actual end is at
base position 181215 of RP11-343N14.
Location/Qualifiers
  1. 181215
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="2"
  /map="2"
  /clone="RP11-343N14"
  /clone_lib="RPC1-11"
  /rpt_family="MERL_type"
  117. 414
  /rpt_family="L1"
  1915. 2634
  /rpt_family="L1"
  2635. 4474
  /rpt_family="L1"
  4475. 4777
  /rpt_family="Alu"
  4778. 4842
  /rpt_family="L1"
  4880. 5125
  /rpt_family="L1"
  5126. 5150
  /rpt_family="(TTG)n"
  5154. 5268
  /rpt_family="Alu"
  5269. 5556
  /rpt_family="L1"
  5554. 6548
  /rpt_family="L1"
  6543. 6609
  /rpt_family="L1"
  6606. 6709
  /rpt_family="L1"
  7833. 7851
  /note="similar to EST AA923166 (NID:g3070475) on18h05.s1"
  7939. 8000
  /rpt_family="AT-rich"
  8936. 9014
  /rpt_family="(TG)n"
  9329. 9357
  /rpt_family="AT-rich"
  9646. 9864
  /rpt_family="MIR"
  10892. 11196
  /rpt_family="Alu"
  10901. 10910
  /note="similar to EST AI288780 (NID:g3932283) qm06f04.x1"
  11878. 11901
  /rpt_family="(TG)n"
  12323. 12372
  /rpt_family="(TG)n"
  <12903. 168586
  /gene="WUGSC:H_NH0343N14.1"
  join(<12903. 12974, 163734. 163846, 168460. 168586)
  /gene="WUGSC:H_NH0343N14.1"
  /note="similar to rsec15 [Rattus norvegicus]; similar to
  AAC01580.1 (PID:g2B27162); H_NH0343N14.1"
  /codon_start=1
  /evidence=not_experimental
  /protein_id="AAF03520.1"
  /db_xref="GI:6094668"
  /translation="FARSGPVGCFQEDTQLQAFIDLRQLDLFIQDWSYLYDGOP
  S"
  NCKLYRVNPTALTLEKMKDTSKNNMFAQFRKNERDKOKLIDTVAKQURGLISSH
  13835. 13846
  misc_feature

```

```

/misc_feature /gene="WUGSC:H_NH0343N14.1"
/notes="similar to EST AA988404 (NID:g3174096) oq89h08.sl"
13835..13846
/repeat_region /gene="WUGSC:H_NH0343N14.1"
/notes="similar to EST A1923304 (NID:g5659268) wn29e05.xl"
16627..16711
/rpt_family="L2"
/repeat_region /rpt_family="L2"
16794..17201
/rpt_family="MarL"
/repeat_region 19331..19425
/rpt_family="L2"
/misc_feature 19677..19698
/gene="WUGSC:H_NH0343N14.1"
/notes="similar to EST A1708260 (NID:g4998036) as24g08.xl"
21103..21181
/rpt_family="Mariner"
/repeat_region 23221..23400
/rpt_family="MIR"
/repeat_region 23938..25320
/rpt_family="L1"
/misc_feature 24835..24870
/gene="WUGSC:H_NH0343N14.1"
/notes="similar to EST AA534098 (NID:g2278114) nj73e04.sl"
25352..25994
/rpt_family="L1"
/repeat_region 26027..26367
/rpt_family="L1"
/repeat_region 26368..26633
/rpt_family="Alu"
/repeat_region 26634..26958
/rpt_family="L1"
/repeat_region 26988..27786
/rpt_family="L1"
/repeat_region 27787..28079
/rpt_family="Alu"
/repeat_region 28080..28717
/rpt_family="L1"
/repeat_region 28789..28814
/rpt_family="AT-rich"
/misc_feature 29010..29023
/gene="WUGSC:H_NH0343N14.1"
/notes="similar to EST AA187741 (NID:g1773942) zp72f12.sl"
29278..29522
/rpt_family="Alu"
/repeat_region 30325..30367
/rpt_family="(TG)n"
/repeat_region 30858..31287
/rpt_family="L2"
/repeat_region 32092..32559
/rpt_family="L2"
/repeat_region 32560..32586
/rpt_family="(CA)n"
/repeat_region 32587..33083

```

```

Query Match 34.3%; Score 26.4; DB 86; Length 181215;
Best Local Similarity 59.2%; Pred. No. 35;
Matches 45; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Oy 1 ctctaaacaaatggttcacgcgttcgctcactcgcggacccgctcaaacgcggccccc 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173456 CTCTGACAGATTTTCAAGCCCTGCTGCTCGCCACCACCTTCCTCCAGGAGCCTCCCCC 173515

Oy 61 ttaacccaacgttaga 76
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173516 TGAACCCCATGTCACA 173531

RESULT 9
ECOW93 338534 bp DNA BCT 30-JAN-2001
LOCUS Escherichia coli K-12 chromosomal region from 92.8 to 00.1 minutes.
DEFINITION U14003
ACCESSION U14003
VERSION U14003.1 GI:1263172

```

```

KEYWORDS Escherichia coli.
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 338534)
AUTHORS Burland,V., Plunkett,G. 3rd, Sofia,H.J., Daniels,D.L. and Blattner,F.R.
TITLE Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes
JOURNAL Nucleic Acids Res. 23 (12), 2105-2119 (1995)
MEDLINE 95334362
REFERENCE 2 (bases 1 to 338534)
AUTHORS Plunkett,G.I.I.I.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1994) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
COMMENT Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
On Apr 17, 1996 this sequence version replaced gi:536929.
This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. The entire sequence was independently determined from E. coli MG1655; overlaps with other sequence determinations are annotated. The start of this entry overlaps the end of the entry ECOW89 (U000006) by 1885 bp.
FEATURES
    Location/Qualifiers
        1..338534
            /organism="Escherichia coli"
            /strain="K-12"
            /sub_strain="MG1655"
            /db_xref="taxon:562"
            /map="92.8 to 100 minutes"
            /note="This sequence comprises the following lambda clones:
                DB928(EC27-278), DB930(EC30-188),
                DB933(EC27-1320), DD935(EC30MM32), DD937(EC27-1070),
                DD941(EC17-142), DD945(EC17-8), DD947(EC24A-34),
                DD949(EC22-169), DD952(EC27-297), DD953(EC27-409),
                DD956(EC17-103), DD958(EC30MM1), DD960(EC30MM2),
                DD962(EC21-104), DD965(EC23A-40), DD968(EC30K660A-4pp),
                DD970(EC19-202), DD974(EC19-61), DD975(EC18-233),
                DD977(EC30MM7), DD980(EC27-1151), DD981(EC18-3),
                DD984(EC17-101), DD987(EC18-115), DD990(EC17-136),
                DD992(EC18-282), DD995(EC17-115), DD997(EC18-113);
                M13mp19 or Janus vectors were used for subcloning"
            1..1885
        /note="1885 bp overlap with end of GenBank Accession Number U00006 (ECOW89)"
        complement(<1..398)
        /note="ORF_f326 of GenBank Accession Number U00006 (ECOW89)"
        /codon_start=1
        /transl_table=11
        /label=ORF_f132p
        /protein_id="AAA96985.1"
        /db_xref="GI:1263173"
        /translation="MGFTRVKSEASEKKPFENALEWDKYGTFFFILAIIVAFCSLSP
        EYFLTNNITQIFVQSSVTVLIGMGEFFAILVAGIDLSVGAILALSGMVTAKMLAGY
        DPFIAAMIGGVLVGGAINGCLVNWITGL"
        277..397
        /note="predicted bend of 75.10 degrees"
        complement(377..1909)
        /note="ORF_f510"
        /codon_start=1
        /transl_table=11
        /label=ORF_f510
        /protein_id="AAA96986.1"
        /db_xref="GI:536931"
        /translation="MATPYISMAGIKSGFVPVHALKSVNLTVYPCETHALLGNCAGK
        STLKMLVSGTHEPTKGTITINNISKLDHKLAAQLGIGIYQELSVDELTVLENLY
        IGRHLTKKICGNLIDWREVRAMLLRLVGLKVDLDERKVAISISHQMLEAKTL
        MLDKAVIIMDEPTSSLTNKEVDYFLIMNQLRKEGTATIVYISHKLAEIRRICDRYTM

```



```

KQSSVCSGIVSDVNDNDIVRLMVGRELONRENAMKENVSNLAHETVFEVRNVTSDR
KKVRDISVSGEILGAGLVSGRTLMCLFGVDRKAGGEIRLNGKDISPRSLD
AVKKMAYITTESRGDFPNFNSIAQNNMAISRLKDGKYGKAGFLFHEYDEQTAENQ
RELLALKHSHVNNQITELSGNQKVLISKWCCCPCEYIIFDEPTRGIDVGAKEIYK
VNRQLADGCKVLWVSSELPEIITVCDRIAVFCESRLTQILTNRDDMSSEETMAMALP
QR"
misc_structure      809.. 931
/feature="predicted bend of 74.75 degrees"
/feature="predicted bend of 75.87 degrees"
/feature="predicted bend of 75.87 degrees"
/feature="promoter matrix score of 55; putative; resembles
flh-regulated promoter"
/standard_name="REP; repetitive extragenic palindromic
element"
/feature="contains 1 REP sequence"
complement(2036..2971)
/feature="ORF_f311"
/codon_start=1
/transl_table=11
/label="ORF_f311"
/protein_id="AAA96987.1"
/db_xref="GI:536932"
/translation="MNKYLKYSGLVGLMLSTSFAAAEYAVVLKTLNPNFVDMKK
GIEDAKTLGVSVDIFAPSGSEDFQSOLFEDLSNKNYKGIAPLSSVNLVMPVAR
AMKGIYLVNDEKIDMNLKAGNVEAFVTDNVAVGAKGASFIDKLGAEGEVA
ITEGKAGNAGSARRGATEAPFKASQIKLVASQPADWDRIKALDVATNVLORNPIK
ATYCANDTMAGVACAVANAGTKGLVVGTDGIPKARKVZGAGTATVAQNPDIG
ATGLKLVDAESKGVIPDLKAPEKLVDSILVTQ"
misc_structure      3010..3121
/feature="predicted bend of 73.22 degrees"
complement(3030..3953)
/feature="ORF_f307"
/codon_start=1
/transl_table=11
/label="ORF_f307"
/protein_id="AAA96988.1"
/db_xref="GI:536933"
/translation="NASSIQOENVMQSOFDSALPNGIGLAPYLRMKOEGMTENESR
IVWLLKQNLSCAPAIKDVAALAVSEAMIVKSKLLGFGFRNLRSALDYFSQSE
OVLPELAFDEAPQDVNVKNFNTIRTHEGQSIYVDEIHRARAFVFOARGDLYGA
GGSNAICADVQHQRIGURVQOAYPDHIMMSASLQEGDVLVTVTHSGRTSDYKAA
VSLAKNGAKICITHSYSHSPIAKLADYICSPAPETPLLRGNASARIQLTLDDAFF
VSAQLNTEQANINMOKTCAIYDFSPGALK"
promoter            complement(3952..3979)
/feature="promoter matrix score of 61; putative"
misc_structure      4151..4251
/feature="predicted bend of 72.06 degrees"
4279..4728
/feature="y1ca"
4279..4728
/feature="y1ca"
/codon_start=1
/transl_table=11
/label="ORF_0149"
/protein_id="AAA96989.1"
/db_xref="GI:536934"
/translation="MKKIAFGCDHVGFILKHEIVAHVLRGVEVIDKGTWSERTDYP
HYASOVALAVAGGEDGILLICGTGVSITRANKFAGIRAVVCSEPYSAQLSRQHNDT
NVLAFGRVGLGELAKMIVDAWLGAQYEGRRHQQRVEAITAEQRN"
misc_feature        complement(4520..16191)
/feature="corresponds to GenBank Accession Number D90227
(ECOPHN) 1..11672; the related sequence GenBank Accession
Number J05260 (ECOPHNAQ) is from E. coli B, not K-12"
complement(4739..5146)
/feature="phnQ"
complement(4739..5146)
/feature="phnQ"
/feature="annotated in GenBank Accession Number J05260
(ECOPHNAQ), but probably not a gene"
/codon_start=1
/transl_table=11

```

```

/protein_id="AAA96990.1"
/db_xref="GI:536935"
/translation="MPAPYTSPIILFHSPLFLIOLPTLLIKLALAIVOLPAALFKL
AAAVILSAPIVAHFTLPAPVARIFIVSPTRVISVINRLNLNANGVVDNIGLRNVD
SVSRVKNRSGKRFISQPKETLSLIIHPCK"
complement(5272..6030)
/feature="phnP"
complement(5272..6030)
/feature="phnP"
/codon_start=1
/transl_table=11
/protein_id="AAA96991.1"
/db_xref="GI:536936"
/translation="MSLTITLTGTGAQGVPAWGCECAACARARRSPQYRPOCPSGVV
KENDATLIDAGLHDLADRWSPGSFOQLLTHYMDHVOCLEFLRWGCVGPPIVYGGP
DSQGDLDLFPKGLLDFSHYVFFVFDLQVYVPLNLHNSKLTFGYLLETAHSRVA
WLSQTAGLPEKTLKFLRNQOPVVMDCSHPPRADAPRNHCDLNTVLALNQVITRSRV
ILTHISHQFDAMLENALPSGEVDFGMEIGVA"
misc_structure      5452..5586
/feature="predicted bend of 75.60 degrees"
complement(6032..6466)
/feature="phnO"
complement(6032..6466)
/feature="phnO"
/codon_start=1
/transl_table=11
/protein_id="AAA96992.1"
/db_xref="GI:536937"
/translation="MPACELRPATQYTDVAYALICELKQAEFDHIAFRVGFNANLRD
PNMRYHALLDGEVGMIGLHLQFHLHVNMIIGEIVVMPOAGRLNVGSKLLAMAE
EPARQAGAEMTSLNVKRDHAFRFLREGYEQSHFRFTKAL"
misc_structure      complement(6453..7010)
/feature="phnN"
complement(6453..7010)
/feature="phnN"
/feature="HisP-like nucleotide binding protein (ATP-binding
protein)"
/codon_start=1
/transl_table=11
/protein_id="AAA96993.1"
/db_xref="GI:536938"
/translation="MNGKLIWLPSPSGKDSLALRELRLREOTQLLVAHRYITRDASA
GSENHALSEQFFTRAGONLLALSHANGLYVGVGEIDLWLHAGFDVLVNGSRHL
POARARYQSALLPVCLOVSPETILRORLENRGRENASEINARLARARYTPQDCHTLNN
DGLSRQSVDTLLTLIHQEKHACL"
misc_structure      complement(7010..8146)
/feature="phnM"
complement(7010..8146)
/feature="phnM"
/feature="HisM-like integral membrane protein"
Query Match      34.3%; Score 26.4; DB 2; Length 338534;
Best Local Similarity 61.8%; Pred. No. 33;
Matches 42; Conservative 0; Mismatches 26; Indels 0; Caps 0;
QY 10 atgttcaagccttcgcttcgctcactcgggaccggcgttaagccggcccttaaccaa 69
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 228713 ATAGTTCTTACCGTTCGGTAATTTTCTCGTGAACCTTTGAATCCGCGCCCTCGGCAAA 228772
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 70 cgttagag 77
| | | | |
Db 228773 TATAACAG 228780
| | | | |
RESULT 10
AL353773/c
LOCUS
DEFINITION
Homo sapiens chromosome 9 clone RP11-574M5, *** SEQUENCING IN
PROGRESS ***, 3 unordered pieces.
ACCESSION
AL353773
VERSION
AL353773.6 GI:13161600
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.

```

```

ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 181931)
Direct Submission
Submitted (24-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Feb 28, 2001 this sequence version replaced gi:10129433.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA574M5
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 180786 bases at least Q40
Consensus quality: 181295 bases at least Q30
Consensus quality: 181617 bases at least Q20
Insert size: 181731; sum-of-contigs
Insert size: 187130; 5.8% error; agarose-fp
Quality coverage: 6.60x in Q20 bases; sum-of-contigs Quality
coverage: 6.75x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7055: contig of 7055 bp in length
* 7056 7155: gap of 100 bp
* 7156 43164: contig of 36009 bp in length
* 43165 43264: gap of 100 bp
* 43265 181931: contig of 138667 bp in length.
FEATURES             Location/Qualifiers
     source           1..181931
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone_lib="RP11-574M5"
                     /clone="RP11-574M5"
                     /clone="RP11-574M5"
misc_feature         1..7055
                     /note="assembly_fragment:02849"
misc_feature         7156..43164
                     /note="assembly_fragment:02176"
                     /fragment_chain:1
misc_feature         43265..181931
                     /note="assembly_fragment:02673"
                     /fragment_chain:1
                     clone_end:SP6
                     vector_side:right
BASE COUNT          52824 a 35727 c 36646 g 56533 t 201 others
ORIGIN
Query Match          33.8%; Score 26; DB 79; Length 181931;
Best Local Similarity 59.5%; Pred. No. 48;
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
QY      3 ctaacaatggttcacgcttcgcttcactcgggaccggcgttaaacggccctt 62
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108165 CCACCAACTGGTAGCTCCCTCCCTTCCTCATTCAGTACCACACATATCATACCCCT 108106
QY      63 aaccaaacgttaga 76
        ||||| |||||

```

```

Db 108105 CTCCAATAATAGA 108092
RESULT 11
AC006239/c
LOCUS      AC006239 195032 bp DNA PRI 31-MAR-1999
DEFINITION Homo sapiens chromosome 9, clone hRPK.467_F_21, complete sequence.
AC006239
ACCESSION AC006239.5 GI:4544480
VERSION    AC006239.5
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195032)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A.,
Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,
McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H.,
Vo,A., Wagner,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (18-DEC-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 195032)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Niloff,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (31-MAR-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 30, 1999 this sequence version replaced gi:4529984.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source           1..195032
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.467_F_21"
                     /clone_lib="RP11-11 human BAC library"
                     /map="9"
                     /chromosome="9"
                     complement(421..576)
repeat_region       /rpt_family="MIR"
                     819..911
repeat_region       /rpt_family="L2"
                     1174..1297
repeat_region       /rpt_family="MIR"

```

```
repeat_region complement(1571..1724)
repeat_region /rpt_family="MER5A"
repeat_region complement(2457..2525)
repeat_region /rpt_family="MIR"
repeat_region 2608..2639
repeat_region /rpt_family="(TTCA)n"
repeat_region complement(2640..2779)
repeat_region /rpt_family="L2"
repeat_region complement(2963..3413)
repeat_region /rpt_family="MLT1E2"
repeat_region complement(3634..4138)
repeat_region /rpt_family="MLT1H"
repeat_region complement(4232..4698)
repeat_region /rpt_family="L1MA9"
repeat_region complement(4699..5019)
repeat_region /rpt_family="MER2"
repeat_region complement(5020..9117)
repeat_region /rpt_family="L1MA9"
repeat_region complement(9118..9428)
repeat_region /rpt_family="AluJo"
repeat_region complement(9429..9603)
repeat_region /rpt_family="L1MA9"
repeat_region complement(9603..10488)
repeat_region /rpt_family="L1M2"
repeat_region complement(11949..12159)
repeat_region /rpt_family="L1M2"
repeat_region complement(12306..12420)
repeat_region /rpt_family="L1M2"
repeat_region 12586..12636
repeat_region /rpt_family="(TTG)n"
repeat_region complement(12735..13104)
repeat_region /rpt_family="HERVL"
repeat_region 13107..13188
repeat_region /rpt_family="MIR"
repeat_region 14010..14091
repeat_region /rpt_family="MIR"
repeat_region complement(14273..15117)
repeat_region /rpt_family="L1ME3A"
repeat_region complement(15118..15474)
repeat_region /rpt_family="MLT1A1"
repeat_region complement(15475..15836)
repeat_region /rpt_family="L1ME3A"
repeat_region complement(15930..16099)
repeat_region /rpt_family="MIR"
repeat_region complement(16421..16490)
repeat_region /rpt_family="MIR"
repeat_region 17393..17423
repeat_region /rpt_family="(CA)n"
repeat_region 18109..18327
repeat_region /rpt_family="MIR"
repeat_region complement(20765..20839)
repeat_region /rpt_family="L2"
repeat_region complement(21217..21364)
repeat_region /rpt_family="MIR"
repeat_region 21588..21619
repeat_region /rpt_family="(CACCAT)n"
repeat_region complement(21793..21958)
repeat_region /rpt_family="MIR"
repeat_region 22369..22374
repeat_region /rpt_family="(T)n"
repeat_region complement(22375..22753)
repeat_region /rpt_family="L1PA7"
repeat_region 22754..22774
repeat_region /rpt_family="(T)n"
repeat_region complement(22775..28038)
repeat_region /rpt_family="L1PA2"
repeat_region complement(28034..2817)
repeat_region /rpt_family="L1PA2"
repeat_region complement(28815..29104)
repeat_region /rpt_family="L1P1"
repeat_region complement(29461..29531)
repeat_region /rpt_family="MIR"
repeat_region 30893..30928
```

```
repeat_region /rpt_family="(CAAAA)n"
repeat_region 31786..31806
repeat_region /rpt_family="AT_rich"
repeat_region complement(33223..33412)
repeat_region /rpt_family="L2"
repeat_region 34813..34995
repeat_region /rpt_family="L1M4"
repeat_region 35075..35115
repeat_region /rpt_family="AT_rich"
repeat_region 35298..35322
repeat_region /rpt_family="(CA)n"
repeat_region 36949..36970
repeat_region /rpt_family="AT_rich"
repeat_region 37155..37374
repeat_region /rpt_family="MIR"
repeat_region 37346..37396
repeat_region /rpt_family="L2"
repeat_region complement(37450..37530)
repeat_region /rpt_family="MIR"
repeat_region complement(37531..37897)
repeat_region /rpt_family="MLT1A1"
repeat_region complement(37898..38035)
repeat_region /rpt_family="MIR"
repeat_region complement(39084..39196)
repeat_region /rpt_family="MIR"
repeat_region 39250..39517
repeat_region /rpt_family="MIR"
repeat_region 39774..40063
repeat_region /rpt_family="AluSg"
repeat_region complement(40557..40909)
repeat_region /rpt_family="THE1A"
repeat_region complement(41349..41511)
repeat_region /rpt_family="L2"
repeat_region 42064..42309
repeat_region /rpt_family="MIR"
repeat_region 42422..42476
repeat_region /rpt_family="AT_rich"
repeat_region 42699..42732
repeat_region /rpt_family="(CAA)n"
repeat_region 42908..43312
repeat_region /rpt_family="MLT1B"
repeat_region 43916..44036
repeat_region /rpt_family="L2"
repeat_region complement(44771..44884)
repeat_region /rpt_family="MIR"
repeat_region complement(44967..45018)
repeat_region /rpt_family="L2"
repeat_region 45218..45804
repeat_region /rpt_family="MER77"
repeat_region complement(46858..46962)
repeat_region /rpt_family="L2"
repeat_region complement(48753..48876)
repeat_region /rpt_family="MIR"
repeat_region 49245..49295
repeat_region /rpt_family="AT_rich"
repeat_region complement(49599..49899)
repeat_region /rpt_family="MLT1J2"
repeat_region 50108..50158
repeat_region /rpt_family="(CATTT)n"
```

Query Match 33.8%; Score 26; DB 86; Length 195032;  
Best Local Similarity 59.5%; Pred. No. 48;  
Matches 44; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 3 ctacaacatggttcgaagccgttcgttcgtcactcggaccggtctaaagccgcccctt 62  
Db 12755 CCACCAACTGGTAGCATCCCTCCCTTCCTCATTCCAGTACCCACACTATCTATACCCCT 12696  
Qy 63 aaccaaacgttaga 76  
Db 12695 CTCCAATAAATAGA 12682



FT	CDS	5954..6298	/codon_start=1	Acinetobacter baumannii integron IntAb74 aminoglycoside 6'-N-acetyltransferase Ib (aacA4) and oxacillinase (oxa-31) genes, complete cds; and unknown gene.
FT		/db_xref="SPTREMBL:Q9MFF4"		
FT		/gene="orf114a"		
FT		/protein_id="BAA99288.1"		
FT		/translation="MRLCTLRIGHLTDPGFRALVGLRDPDFDCLCCVEGISCDRLR		
FT		KARGSKRTGSGIQKRFSSILVLSIRLVLSDFGVSPPFRDELLAPARHFVSVDGGE		
FT		GGLVGKRIVR"		
FT	CDS	complement(6180..6620)		
FT		/codon_start=1		
FT		/db_xref="SPTREMBL:Q9MFF3"		
FT		/gene="orf146"		
FT		/protein_id="BAA99289.1"		
FT		/translation="MCFHAPLKPRKMEKWLFGHTGTAFTVPKIMVTPPLTSPMNFIV		
FT		IVIEIHLVPRONFELAILLPSRQLTSVERMIHSIRIDMRVLHCIRICVIVLKMINH		
FT		LASLIVGSSSRAPLLGPQRQNVGLVPTVHHGRDLSLRDH"		
FT	trna	7075..7146		
FT		/product="trna-Val"		
FT		complement(10981..11301)		
FT	CDS	/codon_start=1		
FT		/db_xref="SPTREMBL:Q9MFF2"		
FT		/gene="orf106a"		
FT		/protein_id="BAA99290.1"		
FT		/translation="MLCYLTGMEDGRLVFAANRIGLSRLSTGLKIRPFRWVWISSL		
FT		RITVLVLAASLTILLQTFILASSPFLRLTRINGFGYLLALPVALQRLVRLFLYEV		
FT		LV"		
FT	CDS	complement(11066..12013)		
FT		/codon_start=1		
FT		/db_xref="SPTREMBL:Q9MFF1"		
FT		/gene="orf315"		
FT		/protein_id="BAA99291.1"		
FT		/translation="MCGKQIKKLYELIRPSAAVHWKRIICNNKSPKSLFITWLGIO		
FT		NRLYTKDRMANNLACDPVCVLCQGAESAGHLFFACEYSQQVWGLFTLRGFPDCCFY		
FT		RLFSMLTGASFTVGQTKVFTYIIPSLCILQKCAVVDLADSLITVNRNDEQII		
FT		ALQPKDRIQELTEFQGVKKLNKSESOLEKRGKAAEQEVVKLEARSQELDAAK		
FT		SATSKATEDRCRSVYADLDFRDYDGGWCAECVREDRFKVIDWAEVDKAFQMDGHIL		
FT		PKDHRAVGCKLPYIVANVHPSELPIPPFENKD"		
FT	CDS	complement(12617..12997)		
FT		/codon_start=1		
FT		/db_xref="SPTREMBL:Q9MFF0"		
FT		/gene="orf126"		
FT		/protein_id="BAA99292.1"		
FT		/translation="MSPFLFAIGMEYLSRCLLELTLPNPNYHPRCEKLAITHMMFASD		
FT		LLIFSADTRSVILLYAQFSSKLPLERRAEYSICEGLEANLDKSSVGGLCFAEKASMQ		
FT		EILGIWMLSPIRSLSSNVSP"		
FT	CDS	complement(12943..13332)		
FT		/codon_start=1		
FT		/db_xref="SPTREMBL:Q9MFF9"		
FT		/gene="orf129a"		
FT		/protein_id="BAA99293.1"		
FT		/translation="MAVIVGEVINEAQAGFIPGKHIGNILLATELIKYSQKAVSPRC		
FT		MIKVDLRKAYDSIEWFLOCMLSLGFPSQFITVIMECVTVSVSILLNGKPTTPTPSK		
FT		KGVQVTPCHPSCLLQWISCLDAY"		
FT	CDS	complement(14842..15243)		
FT		/codon_start=1		
FT		/db_xref="SPTREMBL:Q9MFF8"		
<hr/>				
Query Match 33.5%; Score 25.8; DB 43; Length 199000;				
Best Local Similarity 60.9%; Pred. No. 57;				
Matches 42; Conservative 0; Mismatches 27; Indels 0; Gaps 0;				
Qy	1	ctctaacaatggttcaagcgcttcgctcactcggcgaccggtctaaagcgccgcc 60		
Db	59732	CTATGTCCAATCCTTAACCCATATACCTTTCCTCAGTGGGAATGGATATTTAGGCCCC 59673		
Qy	61	ttacacaaa 69		
Db	59672	TTCACCACA 59664		
<hr/>				
RESULT 13				
AY007784				
LOCUS	AY007784	2051 bp	DNA	BCT 27-NOV-2000
<hr/>				
Query Match 32.7%; Score 25.2; DB 2; Length 2051;				
Best Local Similarity 60.0%; Pred. No. 1.2e+02;				
Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;				
Qy	6	acaaatggttcaagcgcttcgctcactcggcgaccggtctaaagcgccgccctaac 65		





---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2001, 16:53:10 ; Search time 183.37 Seconds  
(without alignments)  
263.666 Million cell updates/sec

Title: US-09-701-626A-78

Perfect score: 77

Sequence: 1 cctcaacaatggttcaagc.....cccttaacacaaacgttagag 77

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_0601.\*  
1: /cgnl\_9/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /cgnl\_9/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /cgnl\_9/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /cgnl\_9/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /cgnl\_9/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /cgnl\_9/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /cgnl\_9/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /cgnl\_9/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /cgnl\_9/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /cgnl\_9/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /cgnl\_9/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /cgnl\_9/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /cgnl\_9/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /cgnl\_9/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /cgnl\_9/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /cgnl\_9/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /cgnl\_9/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /cgnl\_9/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /cgnl\_9/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /cgnl\_9/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /cgnl\_9/gcgdata/3'', seq/geneseq/NA2000.DAT.\*  
22: /cgnl\_9/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77	100.0	77	21	AAZ44967 P. alcaligenes rep
2	74	96.1	74	21	AAZ88516 P. alcaligenes rep
3	74	96.1	77	21	AAZ44966 P. alcaligenes rep
4	74	96.1	14143	21	AAZ44981 P. alcaligenes res
5	71.8	93.2	77	21	AAZ44908 P. alcaligenes rep
6	71.4	92.7	74	21	AAZ88511 P. alcaligenes rep
7	71.4	92.7	77	21	AAZ44952 P. alcaligenes rep
8	71.4	92.7	77	21	AAZ44987 P. alcaligenes rep
9	71.2	92.5	77	21	AAZ44954 P. alcaligenes rep
10	69.8	90.6	77	21	AAZ44900 P. alcaligenes rep
11	69.8	90.6	77	21	AAZ44913 P. alcaligenes rep

12	69.2	89.9	77	21	AAZ44944	P. alcaligenes rep
13	68.6	89.1	77	21	AAZ44911	P. alcaligenes rep
14	68.2	88.6	74	21	AAZ88505	P. alcaligenes rep
15	68.2	88.6	74	21	AAZ88515	P. alcaligenes rep
16	68.2	88.6	74	21	AAZ88520	P. alcaligenes rep
17	68.2	88.6	77	21	AAZ44930	P. alcaligenes rep
18	68.2	88.6	77	21	AAZ44957	P. alcaligenes res
19	68.2	88.6	13382	21	AAZ44984	P. alcaligenes rep
20	67.6	87.8	77	21	AAZ44894	P. alcaligenes rep
21	67.4	87.5	77	21	AAZ44937	P. alcaligenes res
22	67.4	87.5	7300	21	AAZ44982	P. alcaligenes res
23	67.4	87.5	7300	21	AAZ44983	P. alcaligenes rep
24	67	87.0	77	21	AAZ44910	P. alcaligenes rep
25	67	87.0	77	21	AAZ44990	P. alcaligenes rep
26	66.6	86.5	74	21	AAZ88503	P. alcaligenes rep
27	65.4	84.9	77	21	AAZ44964	P. alcaligenes rep
28	65	84.4	77	21	AAZ44915	P. alcaligenes rep
29	65	84.4	77	21	AAZ44929	P. alcaligenes rep
30	64.4	83.6	78	21	AAZ44907	P. alcaligenes rep
31	62.4	81.0	73	21	AAZ88521	P. alcaligenes rep
32	62.4	81.0	78	21	AAZ44933	P. alcaligenes rep
33	61.2	79.5	78	21	AAZ44936	P. alcaligenes rep
34	60.8	79.0	78	21	AAZ44938	P. alcaligenes rep
35	60.4	78.4	78	21	AAZ44912	P. alcaligenes rep
36	60	77.9	78	21	AAZ44963	P. alcaligenes rep
37	59.2	76.9	78	21	AAZ44940	P. alcaligenes rep
38	58.2	75.6	78	21	AAZ44927	P. alcaligenes rep
39	57.2	74.3	75	21	AAZ88508	P. alcaligenes rep
40	57.2	74.3	76	21	AAZ44961	P. alcaligenes rep
41	57.2	74.3	78	21	AAZ44997	P. alcaligenes rep
42	56.2	73.0	78	21	AAZ44948	P. alcaligenes rep
43	56	72.7	78	21	AAZ44956	P. alcaligenes rep
44	55.8	72.5	76	21	AAZ44962	P. alcaligenes rep
45	55.8	72.5	78	21	AAZ44928	P. alcaligenes rep

ALIGNMENTS

RESULT 1  
AAZ44967  
ID AAZ44967 standard; DNA; 77 BP.  
XX  
AC AAZ44967;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes repeat (PAR) element DNA #74.  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; repeat element; PAR; ss.  
XX  
OS Pseudomonas alcaligenes.  
XX  
PN WO9964632-Al.  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
PA (NEWE ) NEW ENGLAND BIOLABS INC.  
XX  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
DR WPI; 2000-116558/10.  
XX  
PT Cloning intact genes used to isolate genes for restriction enzymes  
PS Claim 7a; Page 61; 97pp; English.  
XX

CC This invention describes a novel method for cloning intact,  
CC diversity-selected genes (I) from within gene cassettes (GC) which  
CC comprises identifying DNA repeats that flank GC, hybridizing  
CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
CC fragments containing (I), ligating these fragments into a vector and  
CC transforming cells with the vector. This method is used to clone a wide  
CC variety of prokaryotic genes that provide a selective advantage under  
CC particular conditions, particularly those that encode restriction  
CC enzymes (used as reagents in molecular biology); adhesins (for use in  
CC coating or for targeting molecules or organisms to particular sites,  
CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
CC inhibitors or antagonists of the toxin, or in vaccination, or a  
CC modification methyltransferase. Intact genes can be cloned directly with  
CC a high probability that the orientation of expression is known in advance  
CC and low probability of association with extraneous, possibly toxic,  
CC genes. AA24894-244980 represent the Pseudomonas alcaligenes repeat (PAR)  
CC elements described in the method of the invention.  
XX  
SQ Sequence 77 BP; 19 A; 25 C; 17 G; 16 T; 0 other;

Query Match 100.0%; Score 77; DB 21; Length 77;  
Best Local Similarity 100.0%; Pred. No. 8.4e-20;  
Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ctctcaacaaatggttcaagcgttcgttcgtcactcgtcggaacggtcaagcggcccc 60  
Db 1 ctctcaacaaatggttcaagcgttcgttcgtcactcgtcggaacggtcaagcggcccc 60  
QY 61 ttaacaaacgttagag 77  
Db 61 ttaacaaacgttagag 77

RESULT 2  
AAZ88516  
ID AAZ88516 standard; DNA; 74 BP.  
XX  
AC AAZ88516;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes repeat (PAR) element DNA PARf13.  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; repeat element; PAR; ss.  
XX  
OS Pseudomonas alcaligenes.  
XX  
PN WO9964632-A1.  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
PA (NEW) NEW ENGLAND BIOLABS INC.  
XX  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
DR WPI; 2000-116558/10.

Cloning intact genes used to isolate genes for restriction enzymes -  
Example 1B; Fig 3E; 97pp; English.  
CC This invention describes a novel method for cloning intact,  
CC diversity-selected genes (I) from within gene cassettes (GC) which  
CC comprises identifying DNA repeats that flank GC, hybridizing  
CC oligonucleotides (ON) to these repeats and amplification to produce DNA

CC fragments containing (I), ligating these fragments into a vector and  
CC transforming cells with the vector. This method is used to clone a wide  
CC variety of prokaryotic genes that provide a selective advantage under  
CC particular conditions, particularly those that encode restriction  
CC enzymes (used as reagents in molecular biology); adhesins (for use in  
CC coating or for targeting molecules or organisms to particular sites,  
CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
CC enzymes; toxins that interact with a host, e.g. for synthesis of  
CC inhibitors or antagonists of the toxin, or in vaccination, or a  
CC modification methyltransferase. Intact genes can be cloned directly with  
CC a high probability that the orientation of expression is known in advance  
CC and low probability of association with extraneous, possibly toxic,  
CC genes. AA289504-288521 represent Pseudomonas alcaligenes repeat (PAR)  
CC elements described in the method of the invention.  
XX

SQ Sequence 74 BP; 19 A; 24 C; 16 G; 15 T; 0 other;

Query Match 96.1%; Score 74; DB 21; Length 74;  
Best Local Similarity 100.0%; Pred. No. 1.1e-18;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 ctacaacaaatggttcaagcgttcgttcgtcactcgtcggaacggtcaagcggcccc 62  
Db 1 ctacaacaaatggttcaagcgttcgttcgtcactcgtcggaacggtcaagcggcccc 60  
QY 63 aacaaacgttaga 76  
Db 61 aacaaacgttaga 74

RESULT 3  
AAZ44966  
ID AAZ44966 standard; DNA; 77 BP.  
XX  
AC AAZ44966;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes repeat (PAR) element DNA #73.  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; repeat element; PAR; ss.  
XX  
OS Pseudomonas alcaligenes.  
XX  
PN WO9964632-A1.  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
PA (NEW) NEW ENGLAND BIOLABS INC.  
XX  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
DR WPI; 2000-116558/10.  
XX  
PT Cloning intact genes used to isolate genes for restriction enzymes -  
XX Claim 7a; Page 61; 97pp; English.

CC This invention describes a novel method for cloning intact,  
CC diversity-selected genes (I) from within gene cassettes (GC) which  
CC comprises identifying DNA repeats that flank GC, hybridizing  
CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
CC fragments containing (I), ligating these fragments into a vector and  
CC transforming cells with the vector. This method is used to clone a wide  
CC variety of prokaryotic genes that provide a selective advantage under  
CC particular conditions, particularly those that encode restriction

enzymes (used as reagents in molecular biology); adhesins (for use in coating or for targeting molecules or organisms to particular sites, e.g. for competitive exclusion of a selected pathogen); detoxifying enzymes; toxins that interact with a host, e.g. for synthesis of inhibitors or antagonists of the toxin, or in vaccination, or a modification methyltransferase. Intact genes can be cloned directly with a high probability that the orientation of expression is known in advance and low probability of association with extraneous, possibly toxic, genes. AA24894-244980 represent the Pseudomonas alcaligenes repeat (PAR) elements described in the method of the invention.

Sequence 77 BP; 19 A; 26 C; 17 G; 15 T; 0 other;

Query Match 96.1%; Score 74; DB 21; Length 77;  
Best Local Similarity 100.0%; Pred. No. 1.1e-18;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ctacacaaatggttcaagcgttcgtctcactcgtggaacggcgttaagcgccctt 62  
|||||  
DB 3 ctacacaaatggttcaagcgttcgtctcactcgtggaacggcgttaagcgccctt 62  
|||||

QY 63 aaccaaactgttaga 76  
|||||  
DB 63 aaccaaactgttaga 76  
|||||

RESULT 4  
AA244981  
ID AA244981 standard; DNA; 14143 BP.  
XX  
AC AA244981;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes restriction enzyme PacI superintegron DNA #1.  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; PacI; superintegron; ss.  
XX  
OS Pseudomonas alcaligenes.  
XX  
PN WO9964632-Al.  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
PA (NEWE ) NEW ENGLAND BIOLABS INC.  
XX  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
DR WPI; 2000-116558/10.  
XX  
PT Cloning intact genes used to isolate genes for restriction enzymes -  
XX  
PS Example 1A; Fig 3A-1; 97pp; English.

This invention describes a novel method for cloning intact, diversity-selected genes (I) from within gene cassettes (GC) which comprises identifying DNA repeats that flank GC, hybridizing oligonucleotides (ON) to these repeats and amplification to produce DNA fragments containing (I), ligating these fragments into a vector and transforming cells with the vector. This method is used to clone a wide variety of prokaryotic genes that provide a selective advantage under particular conditions, particularly those that encode restriction enzymes (used as reagents in molecular biology); adhesins (for use in coating or for targeting molecules or organisms to particular sites, e.g. for competitive exclusion of a selected pathogen); detoxifying enzymes; toxins that interact with a host, e.g. for synthesis of inhibitors or

antagonists of the toxin, or in vaccination, or a modification methyltransferase. Intact genes can be cloned directly with a high probability that the orientation of expression is known in advance and low probability of association with extraneous, possibly toxic, genes. This sequence represents a Pseudomonas alcaligenes PacI superintegron DNA sequence.

Sequence 14143 BP; 3936 A; 3446 C; 3311 G; 3449 T; 1 other;

Query Match 96.1%; Score 74; DB 21; Length 14143;  
Best Local Similarity 100.0%; Pred. No. 3.4e-18;  
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ctacacaaatggttcaagcgttcgtctcactcgtggaacggcgttaagcgccctt 62  
|||||  
DB 9095 ctacacaaatggttcaagcgttcgtctcactcgtggaacggcgttaagcgccctt 9154  
|||||

QY 63 aaccaaactgttaga 76  
|||||  
DB 9155 aaccaaactgttaga 9168  
|||||

RESULT 5  
AA244908  
ID AA244908 standard; DNA; 77 BP.  
XX  
AC AA244908;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes repeat (PAR) element DNA #15.  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; repeat element; PAR; ss.  
XX  
OS Pseudomonas alcaligenes.  
XX  
PN WO9964632-Al.  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
PA (NEWE ) NEW ENGLAND BIOLABS INC.  
XX  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
DR WPI; 2000-116558/10.  
XX  
PT Cloning intact genes used to isolate genes for restriction enzymes -  
XX  
PS Claim 7a; Page 59; 97pp; English.

This invention describes a novel method for cloning intact, diversity-selected genes (I) from within gene cassettes (GC) which comprises identifying DNA repeats that flank GC, hybridizing oligonucleotides (ON) to these repeats and amplification to produce DNA fragments containing (I), ligating these fragments into a vector and transforming cells with the vector. This method is used to clone a wide variety of prokaryotic genes that provide a selective advantage under particular conditions, particularly those that encode restriction enzymes (used as reagents in molecular biology); adhesins (for use in coating or for targeting molecules or organisms to particular sites, e.g. for competitive exclusion of a selected pathogen); detoxifying enzymes; toxins that interact with a host, e.g. for synthesis of inhibitors or antagonists of the toxin, or in vaccination, or a modification methyltransferase. Intact genes can be cloned directly with a high probability that the orientation of expression is known in advance and low probability of association with extraneous, possibly toxic,

CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)  
CC elements described in the method of the invention.

XX Sequence 77 BP; 18 A; 26 C; 17 G; 16 T; 0 other;  
SQ

Query Match 93.2%; Score 71.8; DB 21; Length 77;  
Best Local Similarity 97.3%; Pred. No. 7.3e-18;  
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ctctacaaatggttcaagcgttcgtctcgtcactcgtgacgcggtctaaagcgcgcgcc 60  
|||||  
Db 1 ctctacaaatggttcaagcgttcgtctcgtcactcgtgacgcggtctaaagcgcgcgcc 60  
|||||  
QY 61 ttaaccaaacgttag 75  
|||||  
Db 61 ttaaccaaacgttag 75

RESULT 6  
AAZ88511  
ID AAZ88511 standard; DNA; 74 BP.  
XX  
AC AAZ88511;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes repeat (PAR) element DNA PARF8.  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; repeat element; PAR; ss.  
XX  
OS Pseudomonas alcaligenes.  
XX  
PN WO9964632-A1.  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
PA (NEWE ) NEW ENGLAND BIOLABS INC.  
XX  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
DR WPI; 2000-116558/10.  
XX

Cloning intact genes used to isolate genes for restriction enzymes -  
XX Example 1B; Fig 3E; 97pp; English.  
XX  
CC This invention describes a novel method for cloning intact,  
CC diversity-selected genes (I) from within gene cassettes (GC) which  
CC comprises identifying DNA repeats that flank GC, hybridizing  
CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
CC fragments containing (I), ligating these fragments into a vector and  
CC transforming cells with the vector. This method is used to clone a wide  
CC variety of prokaryotic genes that provide a selective advantage under  
CC particular conditions, particularly those that encode restriction  
CC enzymes (used as reagents in molecular biology); adhesins (for use in  
CC coating or for targeting molecules or organisms to particular sites,  
CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
CC enzymes; toxins that interact with a host, e.g. for synthesis of  
CC inhibitors or antagonists of the toxin, or in vaccination, or a  
CC modification methyltransferase. Intact genes can be cloned directly with  
CC and low probability of association with extraneous, possibly toxic,  
CC genes. AAZ88504-288521 represent Pseudomonas alcaligenes repeat (PAR)  
CC elements described in the method of the invention.  
XX  
SQ Sequence 74 BP; 18 A; 23 C; 17 G; 16 T; 0 other;

Query Match 92.7%; Score 71.4; DB 21; Length 74;  
Best Local Similarity 98.6%; Pred. No. 1e-17;  
Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 3 ctacaaatggttcaagcgttcgtctcgtcactcgtgacgcggtctaaagcgcgcctt 62  
|||||  
Db 1 ctacaaatggttcaagcgttcgtctcgtcactcgtgacgcggtctaaagcgcgcctt 60  
|||||  
QY 63 aaccaaacgttag 75  
|||||  
Db 61 aaccaaacgttag 73

RESULT 7  
AAZ44952  
ID AAZ44952 standard; DNA; 77 BP.  
XX  
AC AAZ44952;  
XX  
DT 16-MAY-2000 (first entry)  
XX  
DE P. alcaligenes repeat (PAR) element DNA #59.  
XX  
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;  
KW detoxifying enzyme; repeat element; PAR; ss.  
XX  
OS Pseudomonas alcaligenes.  
XX  
PN WO9964632-A1.  
XX  
PD 16-DEC-1999.  
XX  
PF 11-JUN-1999; 99WO-US13295.  
XX  
PR 12-JUN-1998; 98US-0089086.  
PR 12-JUN-1998; 98US-0089101.  
XX  
PA (NEWE ) NEW ENGLAND BIOLABS INC.  
XX  
PI Raleigh EA, Vaisvila R, Morgan RD;  
XX  
DR WPI; 2000-116558/10.  
XX  
Cloning intact genes used to isolate genes for restriction enzymes -  
XX Claim 7a; Page 60; 97pp; English.

CC This invention describes a novel method for cloning intact,  
CC diversity-selected genes (I) from within gene cassettes (GC) which  
CC comprises identifying DNA repeats that flank GC, hybridizing  
CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
CC fragments containing (I), ligating these fragments into a vector and  
CC transforming cells with the vector. This method is used to clone a wide  
CC variety of prokaryotic genes that provide a selective advantage under  
CC particular conditions, particularly those that encode restriction  
CC enzymes (used as reagents in molecular biology); adhesins (for use in  
CC coating or for targeting molecules or organisms to particular sites,  
CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
CC enzymes; toxins that interact with a host, e.g. for synthesis of  
CC inhibitors or antagonists of the toxin, or in vaccination, or a  
CC modification methyltransferase. Intact genes can be cloned directly with  
CC and high probability of association with extraneous, possibly toxic,  
CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)  
CC elements described in the method of the invention.  
XX  
SQ Sequence 77 BP; 18 A; 25 C; 18 G; 16 T; 0 other;

Query Match 92.7%; Score 71.4; DB 21; Length 77;  
Best Local Similarity 98.6%; Pred. No. 1e-17;

Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctacaaatggttcaagcgttcgctcactcgtgacgcggtctaaagcgcgcctt 62  
 |||||  
 Db 3 ctacaaatggttcaagcgttcgctcactcgtgacgcggtctaaagcgcgcctt 62  
 |||||

QY 63 aaccaaacttag 75  
 |||||  
 Db 63 aaccaaacttag 75  
 |||||

RESULT 8  
 AAZ44987  
 ID AAZ44987 standard; DNA; 77 BP.  
 XX  
 AC AAZ44987;  
 XX  
 DT 16-MAY-2000 (first entry)  
 XX  
 DE P. alcaligenes repeat (PAR) element family 1 consensus DNA #3.  
 XX  
 Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; repeat element; PAR; family 1; ss.  
 XX  
 OS Pseudomonas alcaligenes.  
 XX  
 PN WO9964632-A1.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 11-JUN-1999; 99WO-US13295.  
 XX  
 PR 12-JUN-1998; 98US-0089086.  
 XX  
 PR 12-JUN-1998; 98US-0089101.  
 XX  
 PA (NEW) NEW ENGLAND BIOLABS INC.  
 XX  
 PI Raleigh EA, Vaisvila R, Morgan RD;  
 XX  
 DR WPI; 2000-116558/10.  
 XX  
 PT Cloning intact genes used to isolate genes for restriction enzymes -  
 XX  
 PS Example 1B; Fig 6A; 97pp; English.  
 XX  
 CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites,  
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
 CC enzymes; toxins that interact with a host, e.g. for synthesis of  
 CC inhibitors or antagonists of the toxin, or in vaccination, or a  
 CC modification methyltransferase. Intact genes can be cloned directly with  
 CC a high probability that the orientation of expression is known in advance  
 CC and low probability of association with extraneous, possibly toxic,  
 CC genes. AAZ44985-244987 represents Pseudomonas alcaligenes repeat (PAR)  
 CC element family 1 consensus sequences described in the method of the  
 CC invention.  
 XX  
 SQ Sequence 77 BP; 17 A; 26 C; 18 G; 16 T; 0 other;

Query Match 92.7%; Score 71.4; DB 21; Length 77;  
 Best Local Similarity 98.6%; Pred. No. 1e-17;  
 Matches 72; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ctacaaatggttcaagcgttcgctcactcgtgacgcggtctaaagcgcgcctt 62  
 |||||

Db 3 ctacaaatggttcaagcgttcgctcactcgtgacgcggtctaaagcgcgcctt 62  
 |||||  
 QY 63 aaccaaacttag 75  
 |||||  
 Db 63 aaccaaacttag 75  
 |||||

RESULT 9  
 AAZ44954  
 ID AAZ44954 standard; DNA; 77 BP.  
 XX  
 AC AAZ44954;  
 XX  
 DT 16-MAY-2000 (first entry)  
 XX  
 DE P. alcaligenes repeat (PAR) element DNA #61.  
 XX  
 Diversity-selected gene; restriction enzyme; adhesin; toxin;  
 KW detoxifying enzyme; repeat element; PAR; ss.  
 XX  
 OS Pseudomonas alcaligenes.  
 XX  
 PN WO9964632-A1.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 11-JUN-1999; 99WO-US13295.  
 XX  
 PR 12-JUN-1998; 98US-0089086.  
 XX  
 PR 12-JUN-1998; 98US-0089101.  
 XX  
 PA (NEW) NEW ENGLAND BIOLABS INC.  
 XX  
 PI Raleigh EA, Vaisvila R, Morgan RD;  
 XX  
 DR WPI; 2000-116558/10.  
 XX  
 PT Cloning intact genes used to isolate genes for restriction enzymes -  
 XX  
 PS Claim 7a; Page 60; 97pp; English.  
 XX  
 CC This invention describes a novel method for cloning intact,  
 CC diversity-selected genes (I) from within gene cassettes (GC) which  
 CC comprises identifying DNA repeats that flank GC, hybridizing  
 CC oligonucleotides (ON) to these repeats and amplification to produce DNA  
 CC fragments containing (I), ligating these fragments into a vector and  
 CC transforming cells with the vector. This method is used to clone a wide  
 CC variety of prokaryotic genes that provide a selective advantage under  
 CC particular conditions, particularly those that encode restriction  
 CC enzymes (used as reagents in molecular biology); adhesins (for use in  
 CC coating or for targeting molecules or organisms to particular sites,  
 CC e.g. for competitive exclusion of a selected pathogen); detoxifying  
 CC enzymes; toxins that interact with a host, e.g. for synthesis of  
 CC inhibitors or antagonists of the toxin, or in vaccination, or a  
 CC modification methyltransferase. Intact genes can be cloned directly with  
 CC a high probability that the orientation of expression is known in advance  
 CC and low probability of association with extraneous, possibly toxic,  
 CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)  
 CC elements described in the method of the invention.  
 XX  
 SQ Sequence 77 BP; 18 A; 25 C; 18 G; 16 T; 0 other;

Query Match 92.5%; Score 71.2; DB 21; Length 77;  
 Best Local Similarity 96.1%; Pred. No. 1.2e-17;  
 Matches 73; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tctacaaatggttcaagcgttcgctcactcgtgacgcggtctaaagcgcgcctt 61  
 |||||  
 Db 2 tctacaaatggttcaagcgttcgctcactcgtgacgcggtctaaagcgcgcctt 61  
 |||||

QY 62 taaccaaacttag 77

```
Db 62 taacaaacgttag 77
|||||
RESULT 10
AAZ44900
ID AAZ44900 standard; DNA; 77 BP.
XX
AC AAZ44900;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes repeat (PAR) element DNA #7.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN WO9964632-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
XX
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI Raleigh EA, Vaisvila R, Morgan RD;
XX
DR WPI; 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes
PS Claim 7a; Page 59; 97pp; English.
XX
CC This invention describes a novel method for cloning intact,
CC diversity-selected genes (I) from within gene cassettes (GC) which
CC comprises identifying DNA repeats that flank GC, hybridizing
CC oligonucleotides (ON) to these repeats and amplification to produce DNA
CC fragments containing (I), ligating these fragments into a vector and
CC transforming cells with the vector. This method is used to clone a wide
CC variety of prokaryotic genes that provide a selective advantage under
CC particular conditions, particularly those that encode restriction
CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites,
CC e.g. for competitive exclusion of a selected pathogen); detoxifying
CC enzymes; toxins that interact with a host, e.g. for synthesis of
CC inhibitors or antagonists of the toxin, or in vaccination, or a
CC modification methyltransferase. Intact genes can be cloned directly with
CC a high probability that the orientation of expression is known in advance
CC and low probability of association with extraneous, possibly toxic,
CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
CC elements described in the method of the invention.
XX
SQ Sequence 77 BP; 18 A; 26 C; 17 G; 16 T; 0 other;

Query Match 90.6%; Score 69.8; DB 21; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.1e-17;
Matches 71; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcagcggttcgctcactcggtgacgcggtacaaagccggccctt 62
|||||
Db 3 ctaacaaatggttcagcggttcgctcactcggtgacgcggtacaaagccggccctt 62
QY 63 aaccaaacttag 75
|||||
Db 63 aaccaaacttag 75

RESULT 12
AAZ44944
ID AAZ44944 standard; DNA; 77 BP.
XX
```

```
RESULT 11
AAZ44913
ID AAZ44913 standard; DNA; 77 BP.
XX
AC AAZ44913;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes repeat (PAR) element DNA #20.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN WO9964632-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
XX
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI Raleigh EA, Vaisvila R, Morgan RD;
XX
DR WPI; 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes
PS Claim 7a; Page 59; 97pp; English.
XX
CC This invention describes a novel method for cloning intact,
CC diversity-selected genes (I) from within gene cassettes (GC) which
CC comprises identifying DNA repeats that flank GC, hybridizing
CC oligonucleotides (ON) to these repeats and amplification to produce DNA
CC fragments containing (I), ligating these fragments into a vector and
CC transforming cells with the vector. This method is used to clone a wide
CC variety of prokaryotic genes that provide a selective advantage under
CC particular conditions, particularly those that encode restriction
CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites,
CC e.g. for competitive exclusion of a selected pathogen); detoxifying
CC enzymes; toxins that interact with a host, e.g. for synthesis of
CC inhibitors or antagonists of the toxin, or in vaccination, or a
CC modification methyltransferase. Intact genes can be cloned directly with
CC a high probability that the orientation of expression is known in advance
CC and low probability of association with extraneous, possibly toxic,
CC genes. AAZ44894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
CC elements described in the method of the invention.
XX
SQ Sequence 77 BP; 18 A; 25 C; 17 G; 17 T; 0 other;

Query Match 90.6%; Score 69.8; DB 21; Length 77;
Best Local Similarity 97.3%; Pred. No. 4.1e-17;
Matches 71; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcagcggttcgctcactcggtgacgcggtacaaagccggccctt 62
|||||
Db 3 ctaacaaatggttcagcggttcgctcactcggtgacgcggtacaaagccggccctt 62
QY 63 aaccaaacttag 75
|||||
Db 63 aaccaaacttag 75

RESULT 12
AAZ44944
ID AAZ44944 standard; DNA; 77 BP.
XX
```

```

AC AA244944;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes repeat (PAR) element DNA #51.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN WO9964632-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI Raleigh EA, Vaisvila R, Morgan RD;
XX
DR WPI; 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes
XX
PS Claim 7a; Page 60; 97pp; English.
XX
CC This invention describes a novel method for cloning intact,
CC diversity-selected genes (I) from within gene cassettes (GC) which
CC comprises identifying DNA repeats that flank GC, hybridizing
CC oligonucleotides (ON) to these repeats and amplification to produce DNA
CC fragments containing (I), ligating these fragments into a vector and
CC transforming cells with the vector. This method is used to clone a wide
CC variety of prokaryotic genes that provide a selective advantage under
CC particular conditions, particularly those that encode restriction
CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites,
CC e.g. for competitive exclusion of a selected pathogen); detoxifying
CC inhibitors or antagonists of the toxin, or in vaccination, or a
CC modification methyltransferase. Intact genes can be cloned directly with
CC a high probability that the orientation of expression is known in advance
CC and low probability of association with extraneous, possibly toxic,
CC genes. AA244894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
CC elements described in the method of the invention.
XX
SQ Sequence 77 BP; 16 A; 26 C; 16 G; 19 T; 0 other;

Query Match 89.9%; Score 69.2; DB 21; Length 77;
Best Local Similarity 95.9%; Pred. No. 6.8e-17;
Matches 71; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 tctaacaatggttaacgcgttcgctcactcggaccggtcgaagcgcgcct 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2 tctaactactggttcaagtcgcttcgctcactcggaccggtcgaagcgcgcct 61
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 62 taaccaaacgttag 75
   ||||| |||||
DB 62 taaccaaacgttag 75

RESULT 13
AA244911
ID AA244911 standard; DNA; 77 BP.
XX
AC AA244911;
XX
DT 16-MAY-2000 (first entry)
XX

```

```

DE P. alcaligenes repeat (PAR) element DNA #18.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.
XX
OS Pseudomonas alcaligenes.
XX
PN WO9964632-A1.
XX
PD 16-DEC-1999.
XX
PF 11-JUN-1999; 99WO-US13295.
XX
PR 12-JUN-1998; 98US-0089086.
PR 12-JUN-1998; 98US-0089101.
XX
PA (NEWE ) NEW ENGLAND BIOLABS INC.
XX
PI Raleigh EA, Vaisvila R, Morgan RD;
XX
DR WPI; 2000-116558/10.
XX
PT Cloning intact genes used to isolate genes for restriction enzymes
XX
PS Claim 7a; Page 59; 97pp; English.
XX
CC This invention describes a novel method for cloning intact,
CC diversity-selected genes (I) from within gene cassettes (GC) which
CC comprises identifying DNA repeats that flank GC, hybridizing
CC oligonucleotides (ON) to these repeats and amplification to produce DNA
CC fragments containing (I), ligating these fragments into a vector and
CC transforming cells with the vector. This method is used to clone a wide
CC variety of prokaryotic genes that provide a selective advantage under
CC particular conditions, particularly those that encode restriction
CC enzymes (used as reagents in molecular biology); adhesins (for use in
CC coating or for targeting molecules or organisms to particular sites,
CC e.g. for competitive exclusion of a selected pathogen); detoxifying
CC inhibitors or antagonists of the toxin, or in vaccination, or a
CC modification methyltransferase. Intact genes can be cloned directly with
CC a high probability that the orientation of expression is known in advance
CC and low probability of association with extraneous, possibly toxic,
CC genes. AA244894-244980 represent the Pseudomonas alcaligenes repeat (PAR)
CC elements described in the method of the invention.
XX
SQ Sequence 77 BP; 16 A; 28 C; 18 G; 15 T; 0 other;

Query Match 89.1%; Score 68.6; DB 21; Length 77;
Best Local Similarity 94.7%; Pred. No. 1.1e-16;
Matches 71; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ctctaacaatggttcaacgcgttcgctcactcggaccggtcgaagcgcgcct 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 cctaactactggttcaagtcgcttcgctcactcggaccggtcgaagcgcgcct 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 ttaaccaaacgttag 75
   ||||| |||||
DB 61 ttaaccaaacgttag 75

RESULT 14
AA288505
ID AA288505 standard; DNA; 74 BP.
XX
AC AA288505;
XX
DT 16-MAY-2000 (first entry)
XX
DE P. alcaligenes repeat (PAR) element DNA PARf2.
XX
KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
KW detoxifying enzyme; repeat element; PAR; ss.

```

```
XX OS Pseudomonas alcaligenes.
XX PN WO9964632-A1.
XX PD 16-DEC-1999.
XX PF 11-JUN-1999; 99WO-US13295.
XX PR 12-JUN-1998; 98US-0089086.
XX PA (NEW ) NEW ENGLAND BIOLABS INC.
XX PI Raleigh EA, Vaisvila R, Morgan RD;
XX DR WPI; 2000-116558/10.
XX PT Cloning intact genes used to isolate genes for restriction enzymes -
XX PS Example 1B; Fig 3E; 97pp; English.
XX CC This invention describes a novel method for cloning intact,
XX CC diversity-selected genes (I) from within gene cassettes (GC) which
XX CC comprises identifying DNA repeats that flank GC, hybridizing
XX CC oligonucleotides (ON) to these repeats and amplification to produce DNA
XX CC fragments containing (I), ligating these fragments into a vector and
XX CC transforming cells with the vector. This method is used to clone a wide
XX CC variety of prokaryotic genes that provide a selective advantage under
XX CC particular conditions, particularly those that encode restriction
XX CC enzymes (used as reagents in molecular biology); adhesins (for use in
XX CC coating or for targeting molecules or organisms to particular sites,
XX CC e.g. for competitive exclusion of a selected pathogen); detoxifying
XX CC inhibitors or antagonists of the toxin, or in vaccination, or a
XX CC modification methyltransferase. Intact genes can be cloned directly with
XX CC a high probability that the orientation of expression is known in advance
XX CC and low probability of association with extraneous, possibly toxic,
XX CC genes. AA288504-288521 represent Pseudomonas alcaligenes repeat (PAR)
XX CC elements described in the method of the invention.
XX SQ Sequence 74 BP; 16 A; 25 C; 16 G; 17 T; 0 other;

Query Match 88.6%; Score 68.2; DB 21; Length 74;
Best Local Similarity 95.9%; Pred. No. 1.6e-16;
Matches 70; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcgttcgtcgtcactcgtggaccggtctaaagccggccctt 62
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ctaactactggttcaagtcgttcgtcgtcactcgtggaccggtctaaagccggccctt 60

QY 63 aaccaaacgttag 75
   ||||| |||||
Db 61 aaccaaacgttag 73

RESULT 15
AA288515
ID AA288515 standard; DNA; 74 BP.
XX AC
XX AC AA288515;
XX DT
XX DT 16-MAY-2000 (first entry)
XX DE P. alcaligenes repeat (PAR) element DNA PARf12.
XX KW Diversity-selected gene; restriction enzyme; adhesin; toxin;
XX KW detoxifying enzyme; repeat element; PAR; ss.
XX OS Pseudomonas alcaligenes.
XX PN WO9964632-A1.
```

```
XX PD 16-DEC-1999.
XX PF 11-JUN-1999; 99WO-US13295.
XX PR 12-JUN-1998; 98US-0089086.
XX PA (NEW ) NEW ENGLAND BIOLABS INC.
XX PI Raleigh EA, Vaisvila R, Morgan RD;
XX DR WPI; 2000-116558/10.
XX PT Cloning intact genes used to isolate genes for restriction enzymes -
XX PS Example 1B; Fig 3E; 97pp; English.
XX CC This invention describes a novel method for cloning intact,
XX CC diversity-selected genes (I) from within gene cassettes (GC) which
XX CC comprises identifying DNA repeats that flank GC, hybridizing
XX CC oligonucleotides (ON) to these repeats and amplification to produce DNA
XX CC fragments containing (I), ligating these fragments into a vector and
XX CC transforming cells with the vector. This method is used to clone a wide
XX CC variety of prokaryotic genes that provide a selective advantage under
XX CC particular conditions, particularly those that encode restriction
XX CC enzymes (used as reagents in molecular biology); adhesins (for use in
XX CC coating or for targeting molecules or organisms to particular sites,
XX CC e.g. for competitive exclusion of a selected pathogen); detoxifying
XX CC inhibitors or antagonists of the toxin, or in vaccination, or a
XX CC modification methyltransferase. Intact genes can be cloned directly with
XX CC a high probability that the orientation of expression is known in advance
XX CC and low probability of association with extraneous, possibly toxic,
XX CC genes. AA288504-288521 represent Pseudomonas alcaligenes repeat (PAR)
XX CC elements described in the method of the invention.
XX SQ Sequence 74 BP; 16 A; 24 C; 17 G; 17 T; 0 other;

Query Match 88.6%; Score 68.2; DB 21; Length 74;
Best Local Similarity 95.9%; Pred. No. 1.6e-16;
Matches 70; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ctaacaaatggttcaagcgttcgtcgtcactcgtggaccggtctaaagccggccctt 62
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ctaactactggttcaagtcgttcgtcgtcactcgtggaccggtctaaagccggccctt 60

QY 63 aaccaaacgttag 75
   ||||| |||||
Db 61 aaccaaacgttag 73

Search completed: August 2, 2001, 16:57:28
Job time: 258 sec
```









TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-3903  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3870 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-138-133-1

Query Match 29.1%; Score 22.4; DB 1; Length 3870;  
Best Local Similarity 62.5%; Pred. No. 16;  
Matches 35; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 5 acaaatggttcaagcgttcgctcactcgggacggcgctaaagcgccccc 60  
||||| ||| ||||| || ||| ||||| || ||| || ||||| ||  
Db 391 AACATGTGGATCAAGGCGTTAACTTGCTCATGGAAGATACGCTGCAGCGGCCAC 446

RESULT 5  
US-08-138-641-3  
Sequence 3, Application US/08138641  
Patent No. 5474921  
GENERAL INFORMATION:  
APPLICANT: Koblán, Kenneth S.  
APPLICANT: Pompliano, David L.  
TITLE OF INVENTION: ASSAY TO DETERMINE INHIBITORS OF  
TITLE OF INVENTION: PHOSPHOINOSITIDE-SPECIFIC PHOSPHOLIPASE C-GAMMA  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Muthard  
STREET: P.O. Box 2000, 126 E. Lincoln Avenue  
CITY: Rahway  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/138,641  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muthard, David A.  
REGISTRATION NUMBER: 35,297  
REFERENCE/DOCKET NUMBER: 18937  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-3903  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3893 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-138-641-3

Query Match 29.1%; Score 22.4; DB 1; Length 3893;  
Best Local Similarity 62.5%; Pred. No. 16;  
Matches 35; Conservative 0; Mismatches 21; Indels 0; Gaps 0;  
QY 5 acaaatggttcaagcgttcgctcactcgggacggcgctaaagcgccccc 60  
||||| ||| ||||| || ||| ||||| || ||| || ||||| ||  
Db 391 AACATGTGGATCAAGGCGTTAACTTGCTCATGGAAGATACGCTGCAGCGGCCAC 446

RESULT 6  
US-08-138-133-3  
Sequence 3, Application US/08138133  
Patent No. 5519163  
GENERAL INFORMATION:  
APPLICANT: GIBBS, JACKSON B.  
APPLICANT: KOBLAN, KENNETH S.  
APPLICANT: MACLEOD, ANGUS M.  
APPLICANT: MERCHANT, KEVIN J.  
TITLE OF INVENTION: INHIBITORS OF PHOSPHOINOSITIDE-SPECIFIC  
TITLE OF INVENTION: PHOSPHOLIPASE C  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID A. MUTHARD  
STREET: P.O. BOX 2000, 126 E. LINCOLN AVENUE  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: U.S.A.  
ZIP: 07065  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/138,133  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: MUTHARD, DAVID A.  
REGISTRATION NUMBER: 35,297  
REFERENCE/DOCKET NUMBER: 18938  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 594-3903  
TELEFAX: (908) 594-4720  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3893 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-138-133-3

Query Match 29.1%; Score 22.4; DB 1; Length 3893;  
Best Local Similarity 62.5%; Pred. No. 16;  
Matches 35; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 5 acaaatggttcaagcgttcgctcactcgggacggcgctaaagcgccccc 60  
||||| ||| ||||| || ||| ||||| || ||| || ||||| ||  
Db 391 AACATGTGGATCAAGGCGTTAACTTGCTCATGGAAGATACGCTGCAGCGGCCAC 446

RESULT 7  
US-08-738-172-3  
Sequence 3, Application US/08738172  
Patent No. 5939257  
GENERAL INFORMATION:  
APPLICANT: Szasz, Joseph  
APPLICANT: Davis, Maria  
TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:

```
;; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;; MEDIUM TYPE: storage
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: IBM P.C. DOS 5.0
;; SOFTWARE: Word Perfect 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/738,172
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 60/005,965
;; FILING DATE: October 27, 1995
;; APPLICATION NUMBER: 08/465,003
;; FILING DATE: June 5, 1995
;; APPLICATION NUMBER: 08/575,354
;; FILING DATE: December 20, 1995
;; APPLICATION NUMBER: 08/240,158
;; FILING DATE: May 10, 1994
;; APPLICATION NUMBER: 08/229,329
;; FILING DATE: April 18, 1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Warburg, Richard J.
;; REGISTRATION NUMBER: 32,327
;; REFERENCE/DOCKET NUMBER: 222/158
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (213) 489-1600
;; TELEFAX: (213) 955-0440
;; TELEX: 67-3510
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1368 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-738-172-3

Query Match 28.8%; Score 22.2; DB 2; Length 1368;
Best Local Similarity 61.0%; Pred. No. 15;
Matches 36; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 tctacaataagtgtcaagcgttcgtctcactcgggacgcggtctaaagcggccccc 60
   ||| | |||| | | | | ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 752 TCTTCTGATGGTGGAGGCGCGCATCGACACGCGCGGACATGCCAACGCGCGCC 810

RESULT 8
US-08-738-172-1/c
; Sequence 1, Application US/08738172
; Patent No. 5939257
; GENERAL INFORMATION:
; APPLICANT: Szasz, Joseph
; APPLICANT: Davis, Maria
; TITLE OF INVENTION: THERMOSTABLE ALKALINE PHOSPHATASES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,172
; FILING DATE:
; CLASSIFICATION: 435
```

```
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 60/005,965
;; FILING DATE: October 27, 1995
;; APPLICATION NUMBER: 08/465,003
;; FILING DATE: June 5, 1995
;; APPLICATION NUMBER: 08/575,354
;; FILING DATE: December 20, 1995
;; APPLICATION NUMBER: 08/240,158
;; FILING DATE: May 10, 1994
;; APPLICATION NUMBER: 08/229,329
;; FILING DATE: April 18, 1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Warburg, Richard J.
;; REGISTRATION NUMBER: 32,327
;; REFERENCE/DOCKET NUMBER: 222/158
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (213) 489-1600
;; TELEFAX: (213) 955-0440
;; TELEX: 67-3510
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2936 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-738-172-1

Query Match 28.8%; Score 22.2; DB 2; Length 2936;
Best Local Similarity 61.0%; Pred. No. 18;
Matches 36; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 2 tctacaataagtgtcaagcgttcgtctcactcgggacgcggtctaaagcggccccc 60
   ||| | |||| | | | | ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 791 TCTTCTGATGGTGGAGGCGCGCATCGACACGCGCGGACATGCCAACGCGCGCC 733

RESULT 9
US-07-952-853-21/c
; Sequence 21, Application US/07952853
; Patent No. 5863783
; GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuis, Janna G.
; APPLICANT: Coutel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Flippin, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES
; TITLE OF INVENTION: ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,853
; FILING DATE: 19921125
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
```

```
;
;
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1247..1390
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1391..1442
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1443..1957
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1958..2005
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2006..2089
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2090..2137
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2138..2214
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2215..2262
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2263..2295
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2296..2346
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2347..2498
; FEATURE:
; NAME/KEY: intron
; LOCATION: 2499..2548
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2549..3037
; FEATURE:
; NAME/KEY: intron
; LOCATION: 3038..3092
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3093..3485
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(1247..1390, 1443..1957, 2006..2089,
; LOCATION: 2138..2214, 2263..2295, 2347..2498, 2549..3037,
; LOCATION: 3093..3485)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 1247
; OTHER INFORMATION: /product= "alpha-L-arabinofuranosidase A"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "exoA"
; FEATURE:
```

```
;
; NAME/KEY: sig_peptide
; LOCATION: 1247..1321
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1322..3485
; US-07-952-853-21

Query Match 28.8%; Score 22.2; DB 2; Length 3958;
Best Local Similarity 58.2%; Pred. No. 20;
Matches 39; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 11 tggttcaagccgttcgcttcgctcactcggagaccggtctaaagcggcccttaaccacaaac 70
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 304 TGGTTTACCCGGCTTAATCTCTCCGGGAAGTCGTGCCACCTTCGTATACCCAAC 245

QY 71 gtttagag 77
|||||
Db 244 GTTAGAG 238

RESULT 10
US-08-914-848-21/c
; Sequence 21, Application US/08914848
; Patent No. 5989887
; GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuis, Janna G.
; APPLICANT: Coutel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Flippi, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; TITLE OF INVENTION: ORIGIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914,848
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952,853
; FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
```



```
; Sequence 1, Application US/08832535
; Patent No. 5919658
; GENERAL INFORMATION:
; APPLICANT: NI, JIAN
; APPLICANT: LI, HAODONG
; APPLICANT: YU, GUO-LIANG
; APPLICANT: GENTZ, REINER L
; TITLE OF INVENTION: HUMAN CYSTATIN F
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,535
; FILING DATE: 03-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KIMBALL, PAUL C.
; REGISTRATION NUMBER: 34,610
; REFERENCE/DOCKET NUMBER: PF265
; TELEPHONE: (201) 994-1700
; TELEFAX: (201) 994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 633 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..501
; US-08-832-535-1

Query Match 28.6%; Score 22; DB 2; Length 633;
Best Local Similarity 59.7%; Pred. No. 14;
Matches 37; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 ctctaacaaatggttcacgcggttcgctcactcgagccggtcgaagccgcccc 60
   ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 339 CTCGACCATATATTTCAGGCCCTTCACTATCTGAAC TAGGCCCTTGTGTATGCGGACTC 280

QY 61 tt 62
   |
Db 279 CT 278

RESULT 13
US-09-019-485-1/c
; Sequence 1, Application US/09019485
; Patent No. 6066617
; GENERAL INFORMATION:
; APPLICANT: Li, Haodong
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Gentz, Reiner
; APPLICANT: Ni, Jian
; TITLE OF INVENTION: Cystatin F
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
```

```
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/019,485
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Robert H.
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PF265P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 3013098504
; TELEFAX: 3013098439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 633 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..501
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 67..123
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 124..501
; US-09-019-485-1

Query Match 28.6%; Score 22; DB 3; Length 633;
Best Local Similarity 59.7%; Pred. No. 14;
Matches 37; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 ctctaacaaatggttcacgcggttcgctcactcgagccggtcgaagccgcccc 60
   ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 339 CTCGACCATATATTTCAGGCCCTTCACTATCTGAAC TAGGCCCTTGTGTATGCGGACTC 280

QY 61 tt 62
   |
Db 279 CT 278

RESULT 14
US-08-791-522-2/c
; Sequence 2, Application US/08791522
; Patent No. 5935817
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN CYSTATIN-LIKE
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
```





